

Game for Mayan Speaking Children with Speech Recognition Provided by an English Speech Corpus

Jorge Pech¹, Carlos Miranda², Victor Chi² and Sergio González²

¹ TIC Consultores, Departamento de Desarrollo
Mérida, Yucatán, 97100, México

² Universidad Autónoma de Yucatán, Facultad de Matemáticas, Unidad Multidisciplinaria Tizimín
Tizimín, Yucatán, 97700, México

Abstract

This paper presents a game with speech recognition for children who speak the Mayan language, based on a corpus designed for the English language. The goal of this work is to test whether an English speech corpus can be useful in the development of speech recognition applications for the Mayan language and thus develop teaching resources for the bilingual education.

Keywords: *Speech corpus, speech recognition, Mayan-speaking, Mayan language.*

1. Introduction

Speaking is an effective way of communication available among human beings. It can be defined according to its use as a medium unique, alternative or of support: as a unique medium, is the case where there is no other line of communication; as alternative medium, is when there is other line of communication and if in a certain point that other line were no longer available, then the voice would be used; finally, as a means of support is where in addition to the main line of communication, speech can be used as complement. Likewise, speech is used in all our fields of everyday life either social, working, etc. and is a medium which allow us to convey our ideas, emotions and other feelings.

Since the first attempts in 50s [1] to build machines that work in speech recognition, has been aimed that the human-machine interaction go reaching a greater number of applications, whether for the purposes of research, entertainment, education or others. Currently computer represents a tool of great scope for developing all kinds of activities as well as to become machines that have interaction with humans through the voice. Besides speech being one of the most important communication means in life, it must be emphasized that this medium presents some problems given that there are multiple forms of interpretation of the phonetic sounds we produce, denominated languages. There are more than 6,900 languages around the world and many challenges around

the technologies of spoken language, being the need to support multiple languages of input and output one of those challenges [2]. The Mayan language is considered one of the main indigenous languages of the Americas due to its great influence in the development of the indigenous cultures of Mexico.

After the classic period (250 to 900 AD) there was the popular idea that this language was almost extinct [3], in 2004 there were registered more than 5 million people who speak Maya and in 2009 the INAH warned the possible disappearance over a period of 100 years due to lack of transmission of the Mayan language from parents to children [4].

Currently have been developed and continue in development works on the area of speech recognition, which have been developed for specific languages as for example a continuous corpus for British English recorded at the University of Cambridge [5]; as well as multilingual systems which have several applications in multiple areas. There are systems developed for specific users: the speaker-dependent systems; and those whose purpose is to be used for any talker without additional training: the speaker-independent systems. An example of speaker-independent system is "A Game of Gravity with Speech Recognition for Children with Language Problems" which was developed in the Autonomous University of Yucatan with the purpose of stimulate children affected with language problems [6].

In this paper is presented an Educational Game aimed at children in school age whose dominant language is Mayan, this game is supported by a speech corpus for the English language. This work was done in order to evaluate the contribution of a speech recognition system designed for the English language, when used to recognize the Mayan language. This work is motivated because of mostly of the technological teaching resources already in existence are for English or Spanish languages, and there are few or non-existent for the Mayan language.

2. Framework

2.1 Speech corpus

The idea of a corpus of speech is because humans always need to communicate, to such degree that humans want to communicate even with machines. A speech corpus is a collection of phrases, recorded words of a language in particular [7]. In recent decades have been developed digitalization techniques that have enabled to increase the quality of the recordings. In short, a speech corpus is a database where are stored audio files in standard formats.

Nowadays have been performed many works related to speech corpus: in regards of the language some are for a specific language and some are multilingual; in regards of the dependency on the talker there are speaker dependent and speaker independent works. An example of these is the work presented in [8] in which is described the development of a speech corpus in Spanish for children in school age and with language problems.

2.2 Speech recognition system

The speech recognition system, also known as automatic speech recognition, is the process by which a computer recognizes what a person says .

Despite the great advance in the utilization of speech recognition technologies, by example in mobile devices, there are still challenges which must be worked in, as the linguistic variability, user variability, channel variability, co-articulation to mention some examples [9].

2.3 Bilingual education: indigenous languages

An indigenous language is the language native to a region and spoken in an indigenous town, in which there are still preserved traditions, mores, institutions, or parts of them that are established since before settlers reach those lands [9]. According to the organization Survival International, of approximately 6,000 languages spoken in the world, about 5,000 are indigenous ones [10], so it is inevitable that indigenous languages are in constant interaction with non-indigenous ones. Besides, should be noted that in the interactions of two languages, one of them tend to be dominant, and in the cases where the indigenous language is not the dominant, it is still involved in much of the everyday life aspects such as work, school or socials. These are the cases of our concern. The existence of a bilingual education may become a problem because of the absence of educational models for the educational systems of the countries with these cases. This problem may occur in Latin America since the dominant language is Creole,

due to the contact of the native inhabitants with the settlers who came to this region. Despite this, an example of how to handle bilingual education occurs in the Federal Republic of Germany, where one of its fundamental axes of cooperation is to promote intercultural bilingual education through the planning of educational guidelines ranging from awareness of all actors involved in education unto the formulation of pedagogical aspects [11], this with the idea of improving the quality of education.

2.4 The Mayan language and bilingual education

Even though the work of educating is difficult from the very beginning, the work becomes more complicated when it has to focus on a bilingual intercultural perspective.

The Mayan language has records of existence before the classic period (250 to 900 AD), according to the census conducted in 1990. Table 1 presents the results in the three peninsular states where the Mayan language is spoken [12].

Table 1. Number of mayan-speaking persons and monolingual ones in the states of Yucatan Peninsula

<i>State</i>	<i>Mayan-speakers</i>	<i>Monolingual</i>
Campeche	86,676	5,465
Quintana Roo	133,081	11,114
Yucatan	525,264	40,813
Total	745,021	57,392

In addition to those thrown by the census data, it was obtained that of the 106 municipalities of the state of Yucatán, in which are distributed all Mayan-speaking persons of the peninsula, do not exist municipality in which the indigenous language [12] is not spoken.

2.5 Education and technology

Currently computers have come to occupy an important place in our lives and are found present in nearly every activity we do, as when we go to a store, in cars, at jobs, in our houses and cannot be missed in educational institutions, which play a role as auxiliaries in the teaching activities. Guillermo Cardona says, "This will be the century of knowledge. More precisely the century of scientific and technological rationality" [13], because technology has reached a large part of the activities surrounding us, as in jobs, where employees have to be trained in the technologies that arise. A change of roles of teachers and students is presented along with the presence

of technology in education. Teachers cease to be suppliers of knowledge, now they are facilitators, consultants, motivators and consultants of the student's learning, for which must not only exist a personal interaction, but also virtual interaction. Student ceases to be the passive recipient of knowledge and becomes an autonomous person for his learning [13].

3. Tools for the development of the application

This section describes the CSLU Toolkit, the reasons for which it was used in this work and the advantages that provides.

3.1 CSLU Toolkit

The CSLU Toolkit (Center for Spoken Language Understanding Toolkit) is an integrated set of programs and documents representing the state of the art in tools for research, development and learning on the area of spoken language systems. Its main components are [14]:

- Speech recognition
- Speech synthesis
- Facial animation
- Authoring tools
- Waveforms analysis tools.
- Programming environment

Given that in this work we needed a tool which allows us to work easily with speech recognition and assess the functionality of the corpus, components found in CSLU allowed us to choose it as the best option to develop this work.

3.2 RAD

One of the main advantages of working with the CSLU Toolkit is having an interface named RAD (Rapid Application Developer) allowing us to develop our own applications using its integrated components.

RAD is a graphical tool to create applications focused on structured dialogues which enable the interaction between the user and the computer [15]. It provides a friendly interface that enables the rapid development of applications (Figure 1), has built-in modules to perform specific actions, such as for example a module for voice recognition, another to load images, etc., which are manipulated using the mouse to drag and drop in the workspace.



Figure 1. The RAD interface

Because of the ease provided by RAD, applications can be developed quickly; either for people not related to the area of computation or to make a nicer interface for the end user and thus allow using this tool as a support for teaching courses of speech recognition [16]. Besides using this tool as a teaching resource, can be used to make parts of works such as the presented in the thesis "Design and implementation of a tutorial system based on speech technologies for teaching vocabulary" [7].

4. Development of the application

Among the concerns that give reason to this work can be listed: the null or little existence of technological teaching resources in Mayan language despite the genuine interest on the spread and resurgence of this language; to provide an educational tool for school-age children having Mayan language as their dominant language giving them a training resource which can be used to make their learning easier. Also, the question arises: how to perform tests with children? Because working with children at school age is not as simple as working with a young person or middle-aged, since it is required to get the attention of the participants in the tests of this investigation so to get the expected results. Therefore, it was decided to make a didactic game where the child as well as to help to test the performance of the corpus is comfortable with the use of the tool, so he pay enough attention and most importantly, not having fear to interact with the software.

4.1 Game concept

Since the results expected to be obtained with the application are related to the speech corpus, the game has oral interaction with the participant, so it includes certain

amount of words but in a funny way and not just a simple list of words where the participant repeats them.

The presented application is a didactic game whose purpose is for the participant to discover what animals and objects are hidden behind the silhouettes that cover them. In the corresponding window the participant sees the amount of hidden objects that must be discovered. Using the computer mouse the participant selects any of those objects and using his own voice pronounces the name of the object he believes is behind the silhouette. If the spoken name is correct, its image is displayed. The application uses a system of speech recognition which has a vocabulary of 20 words in the Mayan language, the used corpus is the same included in the CSLU Toolkit system for the English language, and the RAD interface is used for the control of the game. Likewise, for the graphical interface of the game, the modules contained in RAD were used too.

The game is divided into 4 levels, three are in a serial manner, i.e., it has to be finished a level to gain access to the next and so on. It also has an advanced level that is not as intuitive as the others levels and is required some knowledge for the participants to use it. The 20 words that are in Mayan language are divided into the 4 levels of the game.

4.2 Selection of words

Once determined the conceptual idea of the game and the tools to develop it, was selected the right words to be included in the same. For the selection of these were considered: that the words were not unknown and complicated to pronounce for children, in order to avoid complications in the development of the tests; that included a considerable and representative amount of Mayan phonemes, such as *ch'*, *t'*, *k'*, etc.

Because of the above it was decided to use words that fulfil these two conditions, so for the three basic levels were selected names of animals and objects that are known in a popular way in the State of Yucatan, and numbers from one to five were used for the advanced level. The words selected for the game are presented in table 2.

Table 2. Words selected for the game.

Mayan word	English word	Adapted word (phonemes)
Nah	House	Nah
Che'	Tree	Sch eh
K'iin	Sun	Ki in
T'u'ul	Rabbit	Thu uhl
K'éek'en	Pig	keh kehnn
Kaax	Hen	kaash
Am	Spider	Am

Ch'o'	Mouse	Sho o
Tso'	Turkey	Dso o
Xanab	Shoe	shanab
Tsiimin	Horse	Tsiimin
Miis	Cat	Miis
Kaan	Sake	Kaan
Weech	Armadillo	Whehch
Wakax	Cattle	wakash
Hun	One	uhn
Ka'a	Two	Ka a
Oox	Three	Ohsh
Kan	Four	Kan
Ho'	Five	Ho

4.3 Adjustment of words

Once selected the words, it was proceeded to implement them phonetically in the game. In the first instance the words were added just as they were pronounced or written without any changes in its structure, then a series of tests to the word with the CSLU Toolkit voice synthesizer were performed to verify if the pronunciation and phonetic representation were appropriate to the sound that the word in the Mayan language ought to have. If the sound represented by the CSLU Toolkit voice recognition module was not correct then the phonetic structure of the word was modified in an empirical manner, i.e., to trial and error of letter combinations until the word is pronounced as in the Mayan language. At all times the CSLU Toolkit voice synthesizer was used to verify the pronunciation that the word was generating.

4.4 Internal operation of the game

As mentioned previously, the game was developed using modules that by default comes with the RAD of the CSLU Toolkit. The representation of the operation that has each one of the objects in the game can be seen in Figure 2.

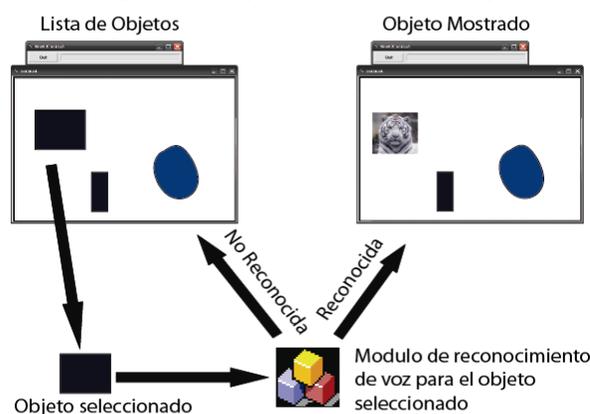


Figure 2. Operation of the game by object

4.5 The game's interface

At the start of the game is presented a window similar to figure 3, to choose one of the levels mentioned above.



Figure 3. The game selection window

From this menu the different parts of the game are accessed. If the image of the number three is selected, is accessed the advanced level where the mechanics is to go counting successively from one to five in Mayan language, for which is deployed a window similar to figure 4. It is started by selecting the number 1 and pronouncing his name. If the pronunciation is correct, then the number two can be selected and then his name can be pronounced, and so on. It should be noted that numbers can only be selected sequentially, from one, and will continue with the next as long as the previous number has been pronounced correctly. When the pronunciation of the number is correct the game paints that number.

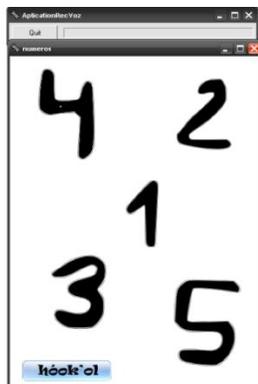


Figure 4. The beginning of the level, with all the numbers unpainted

If another option is selected in the window at the start of the game (Figure 3), it is gained access to its respective level from the three available, where the purpose is to identify the animal that is behind the silhouette. In this section two different ways to play are presented, one in which at the beginning all animals are uncovered (Figure 5) and are being covered one by one as the participant correctly pronounce the name of each animal, and another in which at the beginning all animals are covered (Figure 6).



Figure 5. First play mode, all animals uncovered

At the end of each level, the game emits a voice indicating that all objects have been finished. When all levels are completed the participant will hear the voice "congratulations, you have managed to guess all the objects, congratulations" and the game will return to the main menu allowing another round.

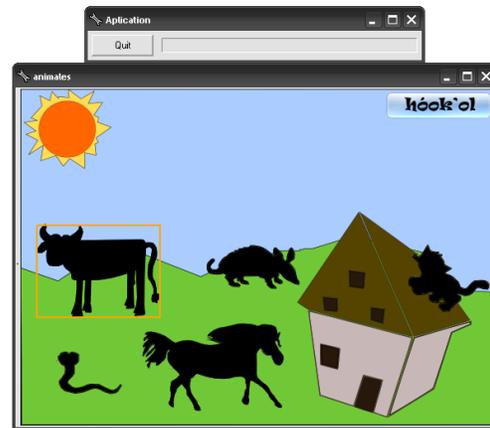


Figure 6. Second play mode, all animals covered

5. Tests and results

In order to test the application it was requested the support of the indigenous primary school "Benito Juárez García" of the city of Tizimin, Yucatan, where we worked with 25 children of first, second and third year with 7, 8 and 9 years old respectively

The objective of the tests was to know if the English corpus could be used in applications that recognize words in the Mayan language. To verify the above it was counted the number of times that the words were recognized correctly by the application.

The interaction of children with the application was carried out in the department of the directorate of the school because it is a place with an environment free of excessive noise, and with adequate space for the computer equipment. For a better control of tests and to have better results, participants were asked to individually go to make use of the game. Taking into account that for the participant may be uncomfortable the presence of a person at his side taking notes, it was decided to make screen recordings during their interaction as well as recording their voices when playing. With these recordings was made the analysis and the counting of the correctly recognized words. Using this technique it was avoided interfering with the tests and our participation was only as auxiliaries in the use of the game.

The results of the interaction of the 25 participants with the game are presented in table 3. This table shows the number of times that the system recognized correctly the pronunciation of a word in the Mayan language regardless of the number of attempts that have failed to recognize.

Table 3. Number of times correctly recognized a word

<i>Word</i>	<i>Number of times recognized</i>
Kaax	24
Kaan	24
Kan	24
Nah	23
Wakax	23
Tsiimin	22
Miis	22
Che'	21
K'iin	21
Am	21
Weech	21
Ka'a	20
K'ëek'en	19
Xanab	19
Hun	19
Ho'	17
T'u'ul	16
Ch'o'	16
Tso'	15
Oox	10

A problem that arose during these tests was the lack of practice of the participants from the elementary school in the use of computers, which complicated their interaction with the game, both in handling the mouse as in the availability of the participants to speak to a microphone. For some participants it was the first time that they had contact with a computer, so before starting with every interaction it was presented to them an explanation of its use and they were motivated to interact verbally with the game, since they showed certain fear to "talk" to the computer. However, once the participants initiated the game, they seemed comfortable with its operation. During the use of the game it was possible to notice the curiosity

and motivation of participants to continue with the game when any word they said was recognized by the system. On the other hand, when a word was not recognized or was not pronounced correctly, the participant became frustrated but tried it again repeating several times. Likewise, when the participant was unaware of any word, asked for it and repeated it several times to pronounce it correctly. Table 3 shows that of the 25 participants that used the game only a word was recognized less than 12 times, as well as 12 words were recognized twenty or more times during the tests. Is worth mentioning that at the time of the analysis of the recordings, some participants showed certain difficulties with the pronunciation of certain words, as for example in the words Ch'o and T'u'ul, which require some accentuation and intonation to be pronounced correctly.

6. Conclusions

This paper presented the development of a game used as tool to verify if a speech corpus of the English language can be used as support for the development of educational tools for the Mayan language, in particular it was used the speech corpus of the CSLU Toolkit. According to the results obtained during the tests, it is concluded that a speech corpus of the English language can be used as a recommended option to develop applications with speech recognition for the Mayan language and in this way to develop applications that are motivating and educational for children whether they speak or not this language.

Also the use of this kind of teaching tools in children at school age would motivate them, in a funny way, to learn. In this work it was successfully developed a didactic tool for the Mayan language, however, and despite the progress in the area of speech recognition, it is still needed to develop more technologies, and that those give support and be extended into more languages, such as the Mayan language.

This work could be the beginning to design a specific speech corpus for the Mayan language, so as to encourage the development of more teaching tools and with them facilitate learning activities to children with a dominant indigenous language. Similarly, these developments can serve as auxiliaries in the recuperation of pronunciation problems of indigenous languages.

References

- [1] Álvarez-Marquina, Historia de los sistemas de reconocimiento automático del habla, Facultad de Informática de la Universidad Politécnica de Madrid, Noviembre, 2001. URL: http://tamarisco.datsi.fi.upm.es/ASIGNATURAS/FRAV/apuntes/historia_rv.pdf

- [2] P. Fung, and T. Schultz, "Multilingual Spoken Language Processing", IEEE Signal Processing Magazine, mayo 2008, pp 89-97. URL: http://csl.anthropomatik.kit.edu/downloads/810_IEEESignal_Processing.pdf
- [3] T. Pérez-Suarez, "Las lenguas mayas: Historia y Diversidad", Revista Digital Universitaria, Vol. 5, No. 7, 2004, pp 2 - 11. URL: http://www.revista.unam.mx/vol.5/num7/art45/ago_art45.pdf
- [4] INAH, "Peligra Lengua Maya", Boletín del INAH, Marzo 2009, Última actualización Mayo 2013. URL: <http://www.inah.gob.mx/index.php/boletines/8-investigaciones-y-estudios-historicos/3039-peligra-lengua-maya>
- [5] T. Robinson, J. Fransen, D. Pye, J. Foote, S. Renals, "WSJCAMO: a British English speech corpus for large vocabulary continuous speech recognition", International Conference on Acoustics, Speech, and Signal Processing, 1995, ICASSP-95 Vol 1, pp.81-84.
- [6] C. Miranda_Palma, R. Camal_Uc, J. Cen-Magaña, C. Gonzalez-Segura, S. Gonzalez-Segura, M. García, and L. Narvaez-Díaz, "Un Juego de Gravedad con Reconocimiento de Voz para Niños con Problemas de Lenguaje", INTERACT 2007, the eleventh IFIP TC13 Internacional Conference on Human-Computer Interaction, 2007.
- [7] T. Luna-Moreno, "Diseño e Implementación de un Sistema Tutorial basado en Tecnologías de Voz para la Enseñanza de Vocabulario", Tesis de Licenciatura, Departamento de ingeniería en sistemas computacionales, Universidad de las Américas Puebla (UDLAP), Puebla, México, 2001.
- [8] G. Perera-Góngora, and C. Miranda Palma, "Diseño de un corpus de Voz en Español para Niños en Edad Escolar con Problemas de Lenguaje", Faz, revista de Diseño de Interacción, Vol. 1, No. 3, 2009, pp. 26-37. URL: http://www.revistafaz.org/n3/disenio_de_corpus.pdf
- [9] M. García-Rojo, "Reconocimiento de voz en medicina", III Congreso Virtual Hispanoamericano de Anatomía Patológica, 2000. URL: <http://www.conganat.org/3congreso/cvhap/conferencias/006/>
- [10] URL: <http://www.elmundo.es/elmundo/2008/02/20/solidaridad/1203525155.html>
- [11] M. Zúñiga, L. Sánchez, D. Zacharías . Demanda y necesidad de educación bilingüe: lenguas indígenas y castellano en el sur andino, 2000, Lima Perú, Informe Final ISBN 9972-854-00-0. URL: http://www.researchgate.net/...Demanda_y_necesidad.../3dec51c3525ad74a7...
- [12] F. Briceño-Chel, Lengua e identidad entre los mayas de la península de Yucatán. URL: <http://www.mayas.uady.mx/articulos/lengua.html>
- [13] G. Cardona-Ossa. Tendencias educativas para el siglo xxi educacion virtual , online y @learning elementos para la discusión. URL: <http://edutec.rediris.es/Revelec2/revelec15/cardona.pdf>
- [14] Universal Speech Tools: The CSLU Toolkit, URL: <http://www.cslu.ogi.edu/toolkit/>
- [15] L. Michel, and R. Zarrougui, CSLU Toolkit: A Technological Survey of a Multimodal Library, 2006. URL: http://diuf.unifr.ch/diva/courses/multimodal/projects/cslu/CLSU_Toolkit.pdf
- [16] B. Serridge, "An Undergraduate Course on Speech Recognition Based on the CSLU Toolkit", International Conference on Acoustics, Speech, and Signal Processing, 1998, ICASSP-98. URL: http://www.mirlab.org/conference_papers/International_Conference/ICSLP%201998/PDF/SCAN/SL980925.PDF

Jorge Rafael Pech-Uicab, hi is a Computer Science graduate for the Universidad Autónoma de Yucatán (UADY). GUI Designer and Software Developer in ERP systems, the main areas of work are Databases (SQL), .NET technology and Web Services.

Carlos Miranda-Palma. Master in Computer Science for the Institute Technology of Monterrey in México, is professor of the Autonomous University of Yucatán. His researcher lines: Voice recognition, Artificial Intelligence, Human Computer Interaction.

Victor Manuel Chi-Pech obtained his degree in Computer Science from the Autonomous University of Yucatan (UADY) in 1996 and his M. Sc. degree in Wireless Network from Monterrey Technological Institute (ITESM), Campus Cuernavaca, in 2007. Victor Chi works since 2000 in the Autonomous University of Yucatan, as a full time professor. He has participated in development projects of software engineering. Currently is giving courses on wireless networks and software engineering in the professional programs in the UADY.

Sergio Gonzalez-Segura. Master in Computer Science for the Centre for Research and Technological Development. Is professor of the Autonomous University of Yucatán and responsible of the electronic laboratory. His researches lines: Intelligent systems, stereo vision, robotics.