Let's Speak Quechua: The Implementation of a Text-to-Speech System for the Incas' Language

John Edgar Vargas Muñoz, Juan Antonio Cruz Tello and Richard Alexander Castro Mamani

Universidad Nacional de San Antonio Abad del Cusco, Av. de la Cultura Nro. 733
Cusco, Perú
jonv147@hotmail.com, j.soccoo@gmail.com, richardtk_1@hotmail.com
WWW home page: http://www.unsaac.edu.pe/

Abstract. The extinction of indigenous languages is one of the most crucial concerns that our societies are facing nowadays. In Peru, there have been many efforts to protect them from disappearance but almost all of them are done without the support of ICT. In this paper, we describe the development process of a text-to-speech (TTS) system for Quechua, the most widely spoken indigenous language of the Americas [1], [2]. For this procedure, we used the Cluster Unit Selection Synthesizer of the Festival Speech Synthesis System. The new male synthetic voice for Quechua will find direct application as part of an assistive technology for visually impaired and illiterate end users. To our knowledge, this project is between the few ones making an attempt to develop a TTS system for an indigenous language in America and the first for Quechua.

Keywords: text-to-speech, system, indigenous, language, Quechua

1 Introduction

The TTS System for Quechua will be an advantage for Quechua speakers, specially for those who do not know how to read, or do not read well. According to table 1 and 2 which describe the illiteracy rate of Quechua speakers, we can observe that more than the 20% of the Quechua speakers do not know how to read or write. And it is just for these people that this system could be useful. This census was done by the INEI (Instituto Nacional de Estadística e Informática, Perú) in 2007.

The Festival Speech Synthesis System is written in C++, has a multi-lingual support, offers a free general framework for building speech synthesis systems for new voices, has a Scheme (SIOD) based command interpreter for control, and has become available for many operating systems like Linux and its many distributions, Mac, and Windows. We chose it for these reasons.

The organization of the paper is as follows. Section 2 discusses Quechua and its phonological aspects. Section 3 discusses the implementation process. Section

Table 1. Numbers of speakers and illiteracy [5], Peru

	1940	1972	1981	2007
Total Pop. Age+	5.227 Mill.	11.331 Mill.	14.573 Mill.	24.687 Mill.
Quechua in %	2,903 Mill.	3,602 Mill.	3,755 Mill.	3,261 Mill.
	55.5%	31.79%	25.77%	13.2%

Table 2. Numbers of speakers and illiteracy [5], Peru

	Total Pop. age 15+	Spanish	Quechua
	19,054,624	15,814,303	2,676,420
literate	17,695,066	15,140,289	2,092,977
illiterate	1,359,558	674,014	583,443
illiteracy rate	7.1%	4.3%	21.8%

4 discusses the results. Then in section 5 we discuss the conclusions.

2 Quechua

Quechua [4] is a group of closely related languages, spoken by 8-10 million people in Peru, Bolivia, Ecuador, Southern Colombia and the North-West of Argentina. Ethnologue also lists some Quechua speakers for Chile. Quechua is one of the official languages of Peru and Bolivia. Peru especially, has increased efforts to provide its citizens with official information not only in Spanish, but also in Quechua and (to less extent) in some other indigenous languages like Aymara and Asháninka.
In fig. 1, we can observe that the Quechuan Languages are divided into two main branches, Quechua I and II in terms of the Peruvian linguist Torero (1964). And these two are subdivided into dialects named used the letters A-C that represent the linguistic distance.

Quechua [4] is an essentially agglutinative language this characteristic differentiates Quechua from the dominant languages in the Americas (mainly English, Spanish and Portuguese), and throughout the years it has struggled with strong social pressure from these dominant languages.

3 Implementation Process with Festival

The first step to build the TTS system is to define the phoneset, for this project we used the Quechua II-C spoken in the city of Cusco with the writing standard proposed by the "Academia Mayor de la Lengua Quechua". The phoneset inventory we used for TTS system was developed according to [6] and [7]. In table 3 and 4 we show these symbols.

Table 3. Vowels and their phonetic representations as they were used in the TTS System for Quechua

GRAPHEME	IPA	SAMPA	VOWEL LENGTH	VOWEL HEIGHT	VOWEL FRONTNESS	LIP ROUNDING
A	a	a	Short	Low	Front	Unrounded
E	ɛ	E	Short	Mid	Front	Unrounded
I	I	I	short	High	Front	Unrounded
O	ɔ	O	Short	Mid	Back	Rounded
U	u	U	Long	High	Back	Rounded

Table 4. Consonants and their phonetic representations as they were used in the TTS System for Quechua

GRAPHEME	IPA	SAMPA	TTS Phonology Symbols	Consonant Type	Place of Articulation	Consonant Voicing
CH	tʃ	tS	ch	Affricative	Palatal	Voiceless
CHH	tʃʰ	tS_h	ch	Affricative	Palatal	Voiceless
CH'	tʃ'	tS_>	chx	Affricative	Palatal	Voiceless
H	h	h	h	Fricative	Post-velar	Voiceless
K	k	k	k	Simple Stop	Velar	Voiceless
KH	kʰ	k_h	kh	Aspirated Stop	Velar	Voiceless
K'	k'	k_>	kx	Glottalized Stop	Velar	Voiceless
L	l	l	l	Lateral Stop	Alveolar	Voiced
LL	ʎ	L	L	Lateral Stop	Palatal	Voiced
M	m	m	m	Nasal Stop	Bilabial	Voiced
N	n	n	n	Nasal Stop	Alveolar	Voiced
Ñ	ɲ	J	N	Nasal Stop	Palatal	Voiced
P	p	p	p	Simple Stop	Bilabial	Voiceless
PH	pʰ	p_h	ph	Aspirated Stop	Bilabial	Voiceless
P'	p'	p_>	px	Glottalized Stop	Bilabial	Voiceless
Q	q	q	q	Simple Stop	Post-velar	Voiceless
QH	qʰ	q_h	qh	Aspirated Stop	Post-velar	Voiceless
Q'	q'	q_>	qx	Glottalized Stop	Post-velar	Voiceless
R	r	r	r	Vibrant Stop	Alveolar	Voiced
S	s	s	s	Fricative Stop	Alveolar	Voiceless
SH	ʃ	S	sh	Fricative	Palatal	Voiceless
T	t	t	t	Simple Stop	Alveolar	Voiceless
TH	tʰ	t_h	th	Aspirated Stop	Alveolar	Voiceless
T'	t'	t_>	tx	Glottalized Stop	Alveolar	Voiceless
W	w	w	w	Approximant	Labial-velar	Voiced
Y	j	j	y	Approximant	Palatal	Voiced

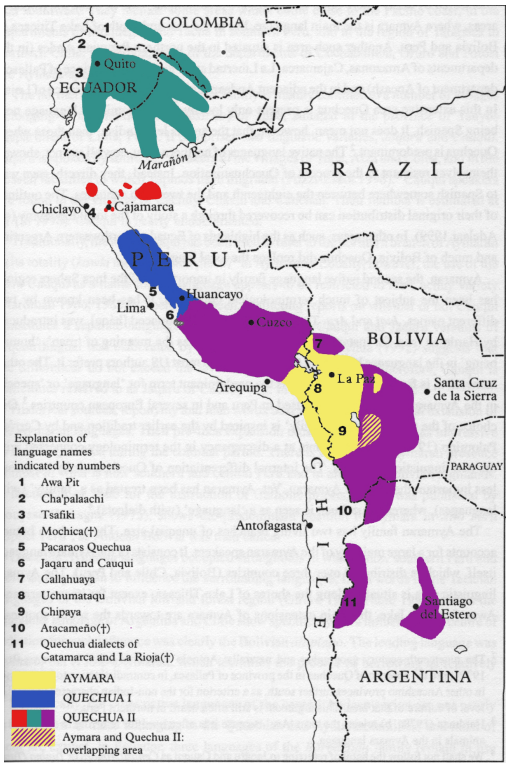

Fig. 1. Map of Quechuan Dialects/ Andean Languages (mid-20th century) (Rios 2010), adapted from Adelaar and Muysken (2004:169).

For the tokenization subprocess which is part of the text analysis, we implemented the rules for the convertion of numbers, Non-Standard words (NSW), to their standard representation, an example of a NSW is given in the table 5.

During the text analysis process, we split the text into tokens using the white space and the punctuation symbols as separators, then each NSW is identified as a separate token by token identifier rules. To identify the tokens we use the scheme regular expression in Festival, After all NSW are identify, we convert them to standard words by means of the pronunciation lexicon or (letter to sound) LTS rules.

The next step is to convert the text into a pronunciation format, for example : allin 'good', is converted to (((a) 1)((ll i n) 0)). Festival provides methods to find the pronunciations of the words. It can consist of three methods: an addenda, typically short, consisting of words added manually; a compiled lexicon, typically large; and a method for dealing with words that are not in either list called LTS Rules (Letter To Sound Rules). In our case, we use the addenda lexicon to define the pronunciation of words that are exceptions in rules. For

Table 5. Non-Standard words

NSW Category	**Written format**	**Pronunciation**
number	10	chunka
number	82	pusaq chunka iskayniyoq
number	5752	pisqa waranqa qanchis pachak pisqa chunka iskayniyoq

example, the lexicon format in festival for the word ari 'yes' is (ari n (((a) 0) ((r i) 1))).
As for the rest of the words we use the LTS rules also called Grapheme to Phoneme (G2P) rules, these rules can be specified in Festival in two ways: by providing Festival with the handcrafted rules or by building the rules automatically, for the Quechua Language we used handcrafted rules because the rules of tranformation from text format to sound format of the language is easier to implement.
The syllable structure of Quechua Language [4] is V, CV, CVC, it is implemented as a rule for a syllabification process in Festival.
For the intonation aspects, we considered that in Quechua, the stress [4] is always on the penultimate syllable, in other words, stress is not fixed, but 'shifts' with the addition of suffixes. There are a few exceptions to this general rule, mainly in connection with special emphatic utterances. These exceptions are marked by accents, e.g. with the suffix -y that expresses emphasis (Dedenbach-Salazar Senz et al. 2002:6).
Ma**ria**qa **as**wan **pi**si mi**khuq**mi **Ma**nuel **man**ta.

Maria -qa as -wan pisi mikhu -q -mi Manuel -manta
Maria -Top bit -Instr little eat -Ag -DirE Manuel -Abl

Maria eats less than Manuel.
With a simple example where one root bears different suffixes and the accent shifts: **All**qo 'dog', All**qo**cha 'little dog', Allqo**cha**lla 'only the little dog', Allqocha**lla**yki 'only your little dog'.

Recording the Voice The text corpus we used in the recording process is not phonetically balanced. For the number of existing written texts in Quechua is not enough or doesn't meet the requirements to develop this procedure. However, the text corpus we used contains all the existing phonemes that variety of Quechua used possesses. The voice for the TTS system comes from a Quechua native speaker originally from Muñani a town in southern Peru, province Azángaro in the region Puno. who has lived in the city of Cusco for 31 years. We should highlight that our speaker has a vast knowledge about the Quechua language spoken in Puno and Cusco (Quechua IIC); The recording process lasted four months approximately, for 2 hours every day in a silent room; we use the best

possible equipment that we find available. We recorded the voice of the speaker using Praat[1], a free software for acoustics analysis, after that with the help of the speaker we chose the recordings that we considered to be optimal, then using GoldWave[2], a professional digital audio editor, we proceeded to clean the recording from any noise and to minimize the volume so that the left outs of the noise disappear.
The labeling process delayed for two months as well and it was made using Praat as well, and example of this is shown in fig. 2.

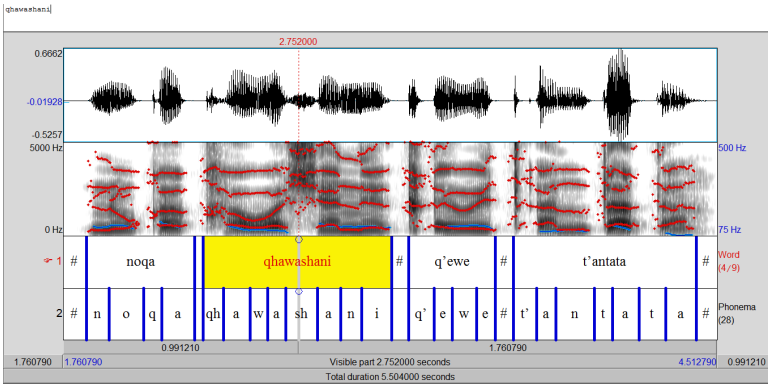

Fig. 2. Praat

Every labeled recording was tested and retested so that we have the best results possible. The emotional aspects of the new voice hasn't been implemented because Qhechua has a extremey complicated system of morphophonemics, this implies that the new voice will have an slightly Quechuan robotic accent. We selected a name for the voice using the convention, we use three parts of names, consisting of the institution name CUS, we employed the name of our city, the language QUZ following the ISO 639-6 standard for the names of languages, and the name of the speaker GUI in this case short for Guillermo.
Name of new voice in Festival: cus_quz_gui

The advantages of using the unit selection method are: first this method produces a more natural sounding speech, and second the unit selection database is created automatically using the Festival framework tools for cluster unit selection, we chose it for these two reasons, the unit selection method is described in the following section.

[1] http://www.fon.hum.uva.nl/praat/
[2] http://www.goldwave.com/

Unit Selection Database for Quechua One of the objectives of the project is to preserve a part of the spoken language in a database and give it a proper use as part of the voice corpus for the TTS system.

Cluster Unit Selection [8] This is a reimplementation of the techniques as described in [9]. The idea is to take a database of general speech and try to cluster each phone type into groups of acoustically similar units based on the (non-acoustic) information available at synthesis time, such as phonetic context, prosodic features (F0 and duration) and higher level features such as stressing, word position, and accents. The actually features used may easily be changed and experimented with as can the definition of the definition of acoustic distance between the units in a cluster.

4 Results

We get to build a working TTS System for Quechua. We have not yet performed any formal evaluation on the system, the system can only read the sentences that have all the vowels and consonants shown before in Table 3 and 4. The limitation of the use of this approach is the amount of manual work that is necessary to develop a TTS system.

5 Conclusions

The specific necessity we want to address with this work is the starting urgency for adaptation of this language to modern communication technologies and as a consequence of this shift we hope that Quechua will survive to the pressure from the dominant languages.
The project have lasted one year and a half now before showing a concrete result, the speaker who gave the voice for the TTS System is a native speaker of the language, having learned Spanish as a second language.
We had to establish a phonetic representation of the Quechua alphabet because there is not an official agreement about this topic yet, and we used this representation successfully in the implementation of the system.
We strongly recommend that researchers in order to obtain a voice that are much more natural do the following:
To increase the amount of recordings, they must be superior to 500 in number, for we have used unit selection method.
There is lots of room for improvement in the system specially in the prosodic aspects of the voice. And we hope that we could produce more TTS systems for the other varieties of Quechua.

6 Acknowledgements

This research was made as part of the undergraduate degree at the faculty of Computer Science of UNSAAC (Universidad Nacional de San Antonio Abad del Cusco), Peru. We would like to thank to Annette Rios for her invaluable insights with the Quechua language, to Magdalena Plamada, Iñaki Sainz, Antonio Bonafonte for their advice on technical issues regarding speech synthesis and the use of the Festival Speech Synthesis System. and to Guillermo Vargas Ochochoque who put the voice for this TTS System.

References

1. Campbell, Lyle: American Indian languages: the historical linguistics of Native America. pg. 253, 1997, Oxford: Oxford University Press.
2. Gordon, Raymond G., Jr.: Ethnologue: Languages of the world (15th ed.), Dallas, Texas: SIL International, 2005, http://www.ethnologue.com
3. Jaén Gomez, Alejandro: Adaptación del Sistema Texto a Voz Festival al Catalán, Departamento de Teoría de la Señal y Comunicaciones, Universidad Politécnica de Cataluña, 2007.
4. Rios, Annette: Applying Finite-State Techniques to a Native American Language: Quechua, Institut für Computer Linguistik, Universitaät Zürich., 2010.
5. Rios, Annette: Sprachtechnologie fr Quechua, Institut für Computer Linguistik, Universitaät Zürich., 2012
6. Cáceres Aragón, Armando: Runasimiq Kurkun, 2007
7. Rodriguez Champi, Albino: Quechua de Cusco, 2006
8. Unit selection databases, http://festvox.org/bsv/c2641.html, Cluster unit selection, www.festvox.org.
9. A. Black and P. Taylor. Automatically clustering similar units for unit selection in speech synthesis, In Eurospeech97, volume 2, pages 601–604, Rhodes, Greece, 1997.
10. Dedenbach-Salazar Sáenz, S., U. von Gleich, R. Hartmann, P. Masson, and C. Soto Ruiz: Rimaykullayki - Unterrichtsmaterialien zum Quechua Ayacuchano, 4. edition, Berlin: Dietrich Reimer Verlag GmbH, 2002.
11. Adelaar, W. F. H. and P. Muysken: *The Languages of the Andes, Cambridge Language Surveys. Cambridge University Press, 2004.