Native Speaker Dependent System for the Development of a Multi-User ASR-Training System for the Mixtec Language

Santiago Omar Caballero Morales and Edgar De Los Santos Ramírez

Technological University of the Mixtec Region, Postgraduate Division, Highway to Acatlima K.m. 2.5, Huajuapan de Leon, Oaxaca, 69000, Mexico
scaballero@mixteco.utm.mx, edgarinteractivo@hotmail.com

Abstract. The Mixtec Language is one of the main native languages in Mexico, and is present mainly in the regions of Oaxaca and Guerrero. Due to urbanization, discrimination, and limited attempts to promote the culture, the native languages are disappearing. Most of the information available about these languages (and their variations) is in written form, and while there is speech data available for listening and pronunciation practicing, a multimedia tool that incorporates both, speech and written representation, could improve the learning of the languages for non-native speakers, thus contributing to their preservation. In this paper we present some advances towards the development of a Multi-User Automatic Speech Recognition (ASR) Training system for one variation of the Mixtec Language that could be used for the design of speech communication, translation, and learning interfaces for both, native and non-native speakers. The methodology and proposed implementation, which consisted of a native Speaker - Dependent (SD) ASR system integrated with an adaptation technique, showed recognition accuracies over 90% and 85% when tested by a male and a female non-native speakers respectively.

Keywords: Speech recognition, native languages, learning interfaces.

1 Introduction

Research on spoken language technology has led to the development of Automatic Speech Recognition (ASR), Text-To-Speech (TTS) synthesis, and dialogue systems. These systems are now used for different applications such as in mobile telephones for voice dialing, GPS navigation, information retrieval, dictation [1, 2, 3], translation [4, 5], and assistance for handicapped people [6, 7].

ASR technology has been used also for language learning, and examples of these can be found in [8, 9, 10] for English, [11] for Spanish and French among others, and [12] for "sign" languages. These interfaces allow the user to practice their pronunciation at home or work without the limitations of a schedule. They also have the advantage of mobility as

some of them can be installed in different computer platforms, or even mobile telephones for basic practicing. However, although there are applications for the most common foreign languages, there are limited (if any) applications for native or ancient languages.

In Mexico there are around 89 native languages still spoken by 6.6 millions of native speakers. Although the number of speakers may be significant (considering the total number of inhabitants in Mexico) this number is decreasing, especially in the Mixtec region.

The population of native speakers of the Mixtec language is decreasing given urban migration and development, culture rejection and limited attempts to preserve the language. This has been expressed by people living in communities in the Mixtec region of Mexico, and this can be corroborated by national statistics that show that the number of people who spoke any native language, 6.3 millions in 2000 (7.1% of the total population) decreased to 6.0 millions in 2005 (6.6% of the total population), and this amount was even higher in 1990 (7.5% of the total population) [13]. This increases the possibility of native languages being lost, as some dialects or variations had less than 10 known speakers (i.e., Ayapaneco, 4 speakers; Chinanteco of Sochiapan, 2 speakers; Mixtec of the Mazateca Region, 6 speakers [13]). In this case, historic antecedents or information about the language is not recorded, making very difficult to recover or save some part of the language. This may happen to other languages with more speakers. The Mixtec Language, with approximately 480,000 speakers, has been reported to lose annually 200 speakers.

To preserve a language is not an easy task, because all characteristics such as grammar rules, written expression, speech articulation, and phonetics must be documented and recorded. Although there are books and dictionaries that among the word definitions include examples about how to pronounce them, this is not as complete as listening the correct pronunciation from a native speaker

We consider that this goal can be accomplished by the use of modern technology such as that used for foreign language learning [10, 11] to promote the language among non-native speakers and thus, to contribute to its preservation. In this work we focus on the development of an ASR-Training system to allow a speaker to practice his or her pronunciation. The methodology and proposed implementation, which consisted of a native Speaker - Dependent (SD) ASR system integrated with a speaker adapting technique, achieved accuracies over 90% and 85% for a male and a female non-native users respectively.

The development of the native ASR-Training system is presented as follows: in Section 2 the details about the phonetics of the reference Mixtec language variation, and the speech corpus developed to build the native ASR system, are shown; in Section 3 the design of the SD native ASR system, which includes the supervised training of the system's acoustic models, and the adaptation technique for its use by non-native users, are shown; in Section 4 the details of the testing methodology by two non-native speakers and the performance of the system in real time are presented and analyzed; finally in Section 5 we discuss about our findings and future work.

2 The Mixtec Language

2.1 Phonetics

The Mixtec Language, or "Tu'un Savi" (Tongue/Language of the Rain) [14], is present mainly in the states of Sinaloa, Jalisco, Guerrero, Puebla, Oaxaca, and Yucatán. With a number of speakers of approximately 480,000, this is one of the main native languages in Mexico. The Mixtec is a tonal language [14], where the meaning of a word relies on its tone, and because of the geographic dispersion of the Mixtec population, there are differences in tones and pronunciations between communities, which in some cases restricts the communication between them [15]. Because of this, each variation of the Mixtec language is identified by the name of a community, for example, Mixtec from Tezoatlán [16], Mixtec from Yosondúa [17], or Mixtec of Xochapa [18], existing significant differences between vocabularies and their meanings: "cat" and "mouse" are respectively referenced as "chító" and "tı́ín" by the Mixtec of Silacayoapan, and as "vilo" and "choto" by the Mixtec of the South East of Nochixtlán. Hence, the Mixtec cannot be considered as a single and homogeneous language, and there is still a debate about its number of variations, which is within the range of 30 [19] to 81[20].

Table 1. Examples of Mixtec words with tones.

Word	**Meaning**	**Word**	**Meaning**
ñoó	*night*	yukú	*who*
ñoo	*town*	yuku	*mountain*
ño̲o	*Palm*	yuku	*leaf*

Table 2. Repertoire of Mixtec phonemes.

No.	**Phoneme**	**No.**	**Phoneme**	**No.**	**Phoneme**
1	/á/	11	/o/	21	/m/
2	/à/	12	/ú/	22	/n/
3	/a/	13	/ù/	23	/nd/
4	/é/	14	/u/	24	/ñ/
5	/e/	15	/ch/	25	/s/
6	/í/	16	/d/	26	/sh/
7	/ì/	17	/dj/	27	/t/
8	/i/	18	/j/	28	/v/
9	/ó/	19	/k/	29	/y/
10	/ò/	20	/l/	30	/sil/

In general, the Mixtec language has three characteristic tones: high, medium, and low [14, 16, 17, 18, 21, 22, 23, 24]. In Table 1 some examples of words that change their meanings based on the tone applied on their vowels are shown, where (_) is used to identify the low tone, (´) the high tone, and the medium tone is left unmarked [14].

Although there are other tone representations, where the low tone also is represented with a horizontal line over the vowel [24], usually the high tone is represented with the diacritical (´).

Based on the phonemes identified in [14, 21-24] and by integrating the different tones in the vowels, the repertoire shown in Table 2 was defined. The low tone is represented by the diacritical (`) while the high tone is represented by (´), the medium tone is unmarked to keep consistency.

The phonetics of the Mixtec has some differences when compared with the Mexican Spanish language. For example, from Table 2:

- The Mixtec phoneme /dj/ represents a sound similar to the Spaniard Spanish **z** (phoneme /z/), which is stronger than **s** (/s/) in both languages. In the phonetics of the Mexican Spanish from the center region both sounds, **z** and **s**, are represented by the phoneme /s/ [25];
- The Mixtec phonemes /sh/ and /ch/ are pronounced in Mexican Spanish as the consonant **X** in the word "**X**icoténcatl" and **CH** in the word "**ch**icle" respectively;
- There are short pauses, uttered as a glottal closure between vowels within a word, which are represented by (') such as in "tu'un" or "ndá'a" ;
- The Mixtec phoneme /n/ sounds as **n** in the Mexican Spanish (associated with the consonant **N**) if it is placed before a vowel, but is mute if placed after the vowel.

2.2 Vocabulary

Because the purpose of the system is to be used for speech training and practicing of the Mixtec language, a vocabulary used for learning was chosen. For this, we established contact with a native speaker who teaches the Mixtec language at the local Cultural Center. The place of origin of this speaker is the community of San Juan Dikiyú in Oaxaca. Since this variation shares similarities with other variations in Oaxaca, we were confident about using it as the reference variation.

With support from the Mixtec teacher we selected 7 traditional Mixtec narratives from a total of 15 that he uses in his lessons for teaching, where the first were used for beginners and the last for more advanced students. The 7 narratives were read twice by the teacher in a recording studio, where the speech samples were recorded in WAV format with a sampling rate of 44,100 Hz and one audio channel (monaural). These recordings were transcribed at the phonetic and word levels (TIMIT standard) using the list of phonemes defined in Table 2 using the software WaveSurfer. All this material formed the **Training Speech Corpus** for the native ASR system which had a total of 192 different words.

Based on the frequency of phonemes of the corpus, which is shown in Figure 1, it was considered that the Training Corpus was phonetically balanced as there were enough samples from each phoneme for the supervised training of the acoustic models of the ASR system.

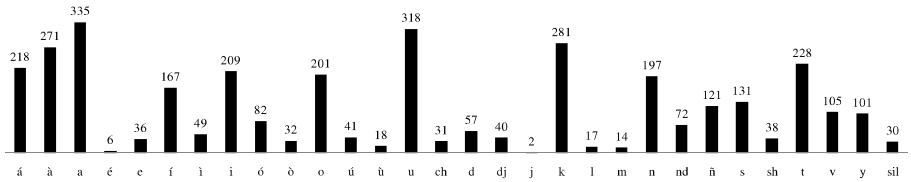

Fig. 1. Frequency distribution of the Mixtec phonemes in the Training Speech Corpus.

3 Mixtec Speaker-Dependent ASR System

The elements of the native ASR that were built with the Training Speech Corpus are shown in Figure 2.

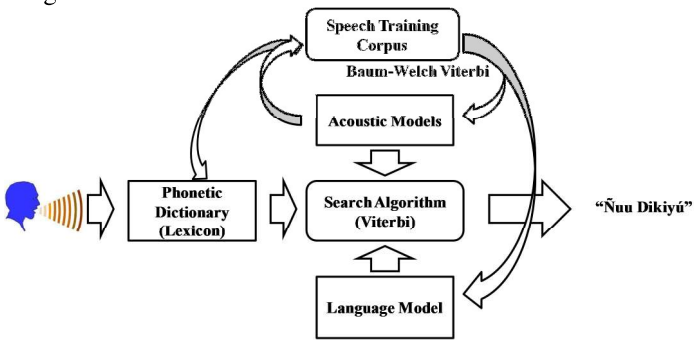

Fig. 2. Structure of the native ASR system.

These were implemented with the software HTK Toolkit [26], and each one was built as follows:

- **Acoustic Models**: Hidden Markov Models (HMMs) [27, 28] were used for the acoustic modeling of each phoneme in the Training Corpus. These were standard three-state left-to-right HMMs with 10 Gaussian components per state. The front-end used 12 MFCCs plus energy, delta, and acceleration coefficients [26]. The supervised training of the HMMs with the Training Corpus (labeled at the phonetic level) was performed with the Baum-Welch and Viterbi algorithms.
- **Lexicon**: the phonetic dictionary was made at the same time as the phonetic labeling of the Training Corpus. The phoneme sequences that formed each word in the vocabulary were defined by perceptual analysis considering the pronouncing rules presented in Section 2.1.

- **Language Model**: Word-bigram language models were estimated from the word transcriptions of the corpus. Speech recognition was performed with a scale grammar factor of 10.
- **Search Algorithm**: Speech recognition was performed with the Viterbi algorithm implemented with the module HVite of the HTK Toolkit.

3.1 Adaptation for Non-native Speakers

As presented in Section 2.2 the Speech Training Corpus of the native ASR was built with the speech samples from a single native speaker. Thus, the system described above is Speaker Dependent (SD) and it will show good performance only when used by the same speaker. For practicing and learning purposes this is a disadvantage.

Commercial ASR systems are trained with hundreds or thousands of speech samples from different speakers, which leads to Speaker-Independent (SI) systems. When a new user wants to use such system, it is common to ask the user to read some words or narratives to provide speech samples that will be used by the system to adapt the SI acoustic models to the patterns of the user's voice. SI ASR systems are robust enough to get benefits by the implementation of adaptation techniques such as MAP or MLLR [26, 28].

In the case of the development of a SI ASR system for the Mixtec language there are challenges given by the wide range of variations in tones and pronunciations, and the limited availability of native speakers to obtain training corpora. Because of this situation, the use of a speaker adaptation technique on this SD system was studied.

Maximum Likelihood Linear Regression (MLLR) was the adaptation technique used for the native SD ASR system in order to make it usable for non-native speakers. MLLR is based on the assumption that a set of linear transformations can be used to reduce the mismatch between an initial HMM model set and the adaptation data. In this work, these transformations were applied to the mean and variance parameters of the Gaussian mixtures of the SD HMMs, and it was performed in two steps:

- **Global Adaptation**. A global base class was used to specify the set of HMM components that share the same transform. Then a *global transform* was generated and applied to every Gaussian component of the SD HMMs.
- **Dynamic Adaptation**. The global transformation was used as an input transformation to adapt the model set, producing better frame/state alignments which were then used to estimate a set of more specific transforms by using a *regression class tree*. For this work, the regression class tree had 32 terminal nodes, and was constructed to cluster together components that were close in acoustic space, and thus could be transformed in similar way. These transforms become more specific to certain groupings of Gaussian components, and are estimated according to the "amount" and "type" of available adaptation data (see Table 3). Because each Gaussian component of an HMM belongs to one particular base class, the tying of

each transformation across a number of mixture components can be used to adapt distributions for which there are no observations at all (hence, all models can be adapted). The adaptation process is dynamically refined when more adaptation data becomes available [26].

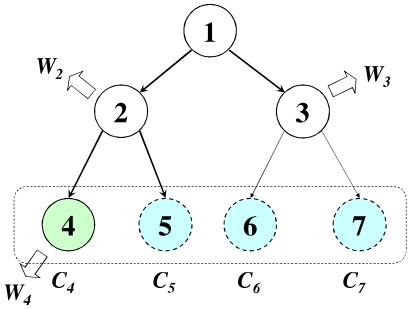

Fig.3. Binary regression class tree with four terminal nodes.

Table 3. Selection of words for supervised non-native speaker adaptation.

No.	Selected Word	Phonemes	No.	Selected Word	Phonemes
1	ÀNE'ECHOOS	à n e e ch o o s	14	KOÚNI	k o ú n i
2	ÁTOKÓ	á t o k ó	15	KUALÍ	k u a l í
3	DIKIYÚ	d i k i y ú	16	KÙTAKU	k ù t a k u
4	DJAMA	d j a m a	17	KUTÓ	k u t ó
5	DJÀVÌ	z à v ì	18	LAA	l a a
6	CHÁNÍ	ch á n í	19	LULI	l u l i
7	CHI	ch i	20	NDAKONÓ	nd a k o n ó
8	ÍDJONA	í z o n a	21	NDEEYÉ	nd e y é
9	ÍÑÒ	í ñ ò	22	NÍKÉE	n í k é e
10	KAMA	k a m a	23	ÑA	ñ a
11	KÌVÌ	k ì v ì	24	ÑUU	ñ u u
12	KOKUMI	k o k u m i	25	SÁXI	s á sh i
13	KÒÒÍÚN	k ò í ú	26	ÙXÌ	ù sh ì

As an example of how MLLR works, in Figure 3 a regression class tree is presented with four terminal nodes (or base classes) denoted as C_4, C_5, C_6 and C_7. Solid nodes and arrows indicate that there is enough data in that class to generate a transformation matrix, and dotted lines and circles indicate that there is insufficient data. During the "dynamic" adaptation, the mixture components of the models that belong to the nodes **2**, **3** and **4** are

used to construct a set of transforms denoted by W_2, W_3 and W_4. When the transformed model set is required, the transformation matrices (mean and variance) are applied in the following fashion to the Gaussian components in each base class: $W_2 \rightarrow C_5$; $W_3 \rightarrow \{C_6, C_7\}$; and $W_4 \rightarrow C_4$, thus adapting the distributions of the classes with insufficient data (nodes **5**, **6**, and **7**) as well as the classes with enough data.

For the native SD system, a selection of words from the Training Corpus was defined to allow the user to provide enough speech samples (adaptation data) from each phoneme listed in Table 2. These words are shown in Table 3 and have the frequency distribution of phonemes shown in Figure 4, which has a correlation coefficient of 0.69 with the distribution of the Training Corpus (Figure 1). Hence it was considered that the adaptation samples were representative of the Training Corpus and sufficient for MLLR adaptation.

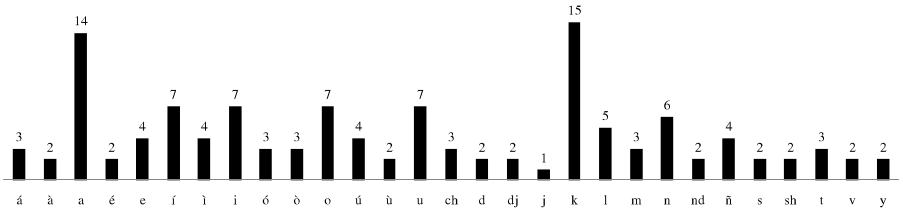

Fig. 4. Frequency distribution of the phonemes in the selection of words for non-native speaker adaptation.

3.2 Non-native Speakers

Two non-native speakers, a female and a male, were recruited to test the performance of the native SD ASR with the adaptation technique. Their background details are shown in Figure 5.

Fig. 5. Non-native speakers for evaluation of the native SD ASR system.

Prior to use the system, both received 6 hours of informative sessions which were distributed over three days. In these sessions, information about the pronunciation of the Mixtec words from the ASR system's vocabulary and the 7 narratives, which included the audios from the native speaker, were reviewed.

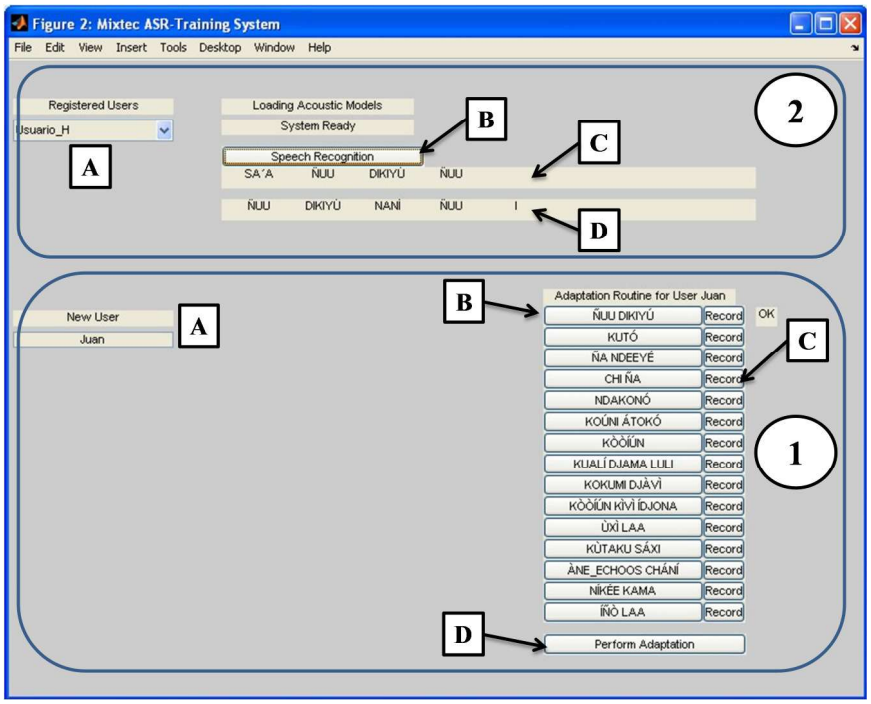

Fig. 6. Graphical User Interface for the native SD ASR-Training system.

3.3 Graphical User Interface

As shown in Figure 6, the native SD ASR system was integrated with a Graphical User Interface (GUI) for the adaptation and recognition tasks which were conducted as follows:

– **Adaptation (1).** As shown in Figure 6, there is a "New User" field (1A) where the user can write his/her name, for example, "Juan". When the user does this, the interface builds the respective directories and files to perform adaptation. On the right side there are buttons with the names of the adaptation words from Table 3 (1B), and with the label "Record" (1C). When the 1B buttons are pressed the audio file corresponding to that word (from the native speaker) is played, so the user can hear the correct pronunciation of that word before providing any speech sample for adaptation. When pressed the respective 1C button (the one next to that word) the interface records the user's pronunciation of that word. When the recording task is finished an "OK" is shown next to 1C. After all adaptation words are recorded the user can press the "Perform Adaptation" button (1D), which starts the MLLR adaptation with the audio samples from the user.

– **Recognition (2).** Once that the user got registered and performed adaptation, his/her data (i.e., MLLR transformations) is stored in directories identified by his/her name. After re-starting the interface, the new user's name is shown in the list of "Registered Users" (2A). At this point the user selects his/her name and the interface automatically loads the corresponding MLLR transformations and acoustic models, enabling the button "Speech Recognition" (2B) to perform ASR in real time when pressed. The user pronounces any phrase from the narratives (when pressing 2B) and the interface displays two outputs: in the field 2C the non-adapted response of the SD ASR is shown, while in 2D the MLLR adapted response is shown.

4 Performance of the Mixtec SD ASR-Training System

The measure of performance for the Mixtec ASR-Training system was the Word Recognition Accuracy (WAcc), which is analogous to the Word Error Rate (WER) [26]. For convenience we used both measures, which are defined as:

$$WAcc = (N-D-S-I)/N \; . \tag{1}$$

$$WER = 1 - WAcc \; . \tag{2}$$

where N is the total number of elements (words) in the reference (correct) transcription of the spoken words, D and I are the number of elements deleted and inserted in the decoded sequence of words (word output from the ASR system), and S the number of elements from the correct transcription substituted by a different word in the decoded sequence.

The Mixtec ASR was tested initially with the Training Corpus, and the performance results are shown in Table 4.

Table 4. Performance of the Mixtec ASR system when tested with the Training Corpus.

N	D	S	I	%WAcc	%WER
911	0	18	29	94.84	5.16

By replacing the word-bigram language model with a phoneme-based language model, a response at the phonetic level was obtained from the recognizer. A phoneme confusion-matrix, shown in Figure 7, was estimated from this response in order to identify patterns of errors at the low level of the baseline ASR.

As it can be observed, there were a few confusions between phonemes, for example: between vowels /á/, /à/, /a/, and /í/, /ì/, /i/; and a significant confusion between /nd/ and /d/. Analogous to Table 4, and as presented in Figure 7, the performance of the Mixtec ASR at the phonetic level is shown in Table 5.

As shown in Figure 7 and Table 5, the deletions and substitutions rates were approximately 10% of N (most of the deleted phonemes were vowels), while insertions represented approximately 5%. A %WAcc of 78% is normal based on the fact that there

was no restriction from the phonetic dictionary (Lexicon) to form valid sequences of phonemes (which lead to a %WAcc of 94.84%). These results show that the acoustic modeling of the tonal phonemes of the SD ASR system with the Training Corpus was performed satisfactorily. This is normal in most cases unless there were many variations or inconsistencies in the training speech.

Table 5. Performance of the phoneme-based Mixtec ASR system when tested with the Training Corpus.

N	D	S	I	%WAcc	%WER
3846	340	338	144	78.63	21.37

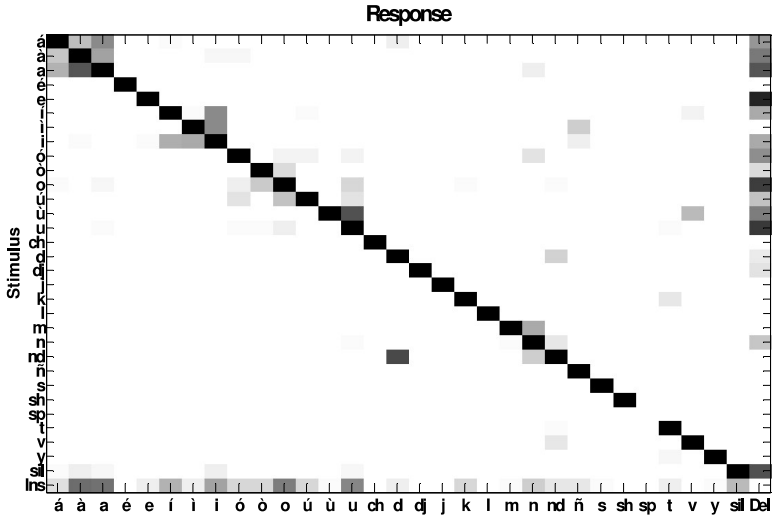

Fig. 7. Pattern of errors at the phonetic level of the Mixtec ASR system when tested with the Training Corpus.

The system was tested by the non-native speakers using three narratives (NTVs) with different levels of difficulty: 1 (easy level), 3 (medium level), and 6 (hard level). Each user was allowed to try up to 10 times the system in case that his/her uttered phrase wasn't recognized. If after those trials the phrase was not recognized, then the last result was recorded as the final response of the ASR system.

The performance results for the non-native speakers are shown in Table 6. The accuracy for the male user was -4.46% when no adaptation was performed. After the adaptation session, the performance increased to 92.57%. With the female user the non-adapted system performed with an accuracy of 9.90%, however after the adaptation session this increased to 88.12%.

Table 6. Performance of the Non-adapted and Adapted Mixtec ASR system when tested by two non-native speakers.

User GCV							User SCB						
Non-adapted SD ASR-Training System							Non-adapted SD ASR-Training System						
NTV	N	D	S	I	%WAcc	%WER	NTV	N	D	S	I	%WAcc	%WER
1	53	4	32	13	7.55	92.45	1	53	8	30	7	15.09	84.91
3	53	2	32	20	-1.89	101.89	3	53	6	25	18	7.55	92.45
6	96	1	59	48	-12.50	112.50	6	96	4	66	18	8.33	91.67
Total	202	7	123	81	-4.46	104.46	Total	202	18	121	43	9.90	90.10
MLLR-adapted SD ASR-Training System							MLLR-adapted SD ASR-Training System						
NTV	N	D	S	I	%WAcc	%WER	NTV	N	D	S	I	%WAcc	%WER
1	53	0	1	0	98.11	1.89	1	53	0	2	0	96.23	3.77
3	53	0	3	2	90.57	9.43	3	53	0	5	3	84.91	15.09
6	96	0	8	1	90.63	9.38	6	96	0	10	4	85.42	14.58
Total	202	0	12	3	92.57	7.43	Total	202	0	17	7	88.12	11.88

While performing the testing sessions there were recorded the number of trials before the adapted system could recognize correctly the test phrase. For the male user (GCV), in 7 out of 49 phrases after the 10 trials the exact phrase was not obtained. A mean of 2.97 with a standard deviation of 3.17 was obtained for the number of trials before the recognizer could decode the correct phrase. In contrast, for the female user, in 17 phrases the correct phrase was not obtained after the 10 trials. This was reflected in a mean of 5.10 trials with a standard deviation of 4.08 for this user.

These differences in performances could be caused by the acoustic differences between the female's voice and the male's voice. Also, variations in the level of knowledge or ability to utter the Mixtec phonemes, which are user dependent, can be attributable factors. A matched pairs test [29] was used to test for statistical significant differences between both performances, obtaining a p-value of 0.21 (< 0.10) for the results presented in Table 6. Because of this, it was concluded that there was no statistical difference between both performances.

5 Conclusions and Future Work

In this paper we presented our advances towards the development of a Multi-user Mixtec ASR system for language learning purposes. The Mixtec language is a complex language given the diversity of tones and vocabulary, and so there are challenges to accomplish such system, especially with limited availability of native speakers for speech corpora. Nevertheless a native SD ASR system was developed for purposes of pronunciation training of a tonal language. This system, when integrated with a speaker adaptation technique, performed with levels of recognition accuracy of 92.57% (male user) and 88.12% (female user) for two non-native speakers, thus making it a multi-user system.

As presented in Table 6, MLLR is a very reliable adaptation technique when applied on a tonal SD ASR system, being able to accomplish with few speech samples (in this case, 26 words) improvements in recognition accuracy of around 90%. For non-native users, most of the word recognition errors were substitutions (12 for male user, 17 for female user) and insertions (3 and 7 respectively). As starting point, the results presented in this paper are encouraging, however we do realize that much more research is needed, and here we present our future work:

- Develop a technique to increase the performance of the native ASR system.
- Increase the Training Speech Corpus: add more vocabulary words and increase the complexity of the narratives; recruit more native speakers (both genders) in order to develop a native SI ASR system. Currently we are in talks to recruit three additional native speakers.
- Test the system with more users with different levels of expertise in the Mixtec language (group tests are being planned).
- Improve the GUI to increase usability: incorporate learning methodologies to extend the use of the ASR system for users that don't have previous knowledge of the language (with no informative sessions); integrate a measure of performance for the level of knowledge or practicing that the user gets by using the ASR system.

References

1. Philips: SpeechExec Pro Transcribe. http://www.dictation.philips.com/index.php?id=1440&CC=VV
2. Nuance: Dragon Speech Recognition Software. http://www.nuance.com/dragon/index.htm
3. IBM: Embedded ViaVoice.http://www-01.ibm.com/software/pervasive/embedded_viavoice/
4. Carnegie Mellon University (The Interactive Systems Laboratories): JANUS Speech Translation System. http://www.is.cs.cmu.edu/mie/janus.html
5. The German Research Center for Artificial Intelligence: Verbmobil - Translation of Spontaneous Speech. http://www.dfki.de/lt/project.php?id=Project_382&l=en
6. Green, P.D., Hawley, M.S., Enderby, P., Cunningham, S.P., Parker,M.: Automatic speech recognition and training for severely dysarthric users of assistive technology: The STARDUST project. Clinical Linguistics and Phonetics, 20:149–156 (2006)
7. Hawley, M., Cunningham, S., Cardinaux, F., Coy, A., O'Neill, P., Seghal, S., Enderby, P.: Challenges in developing a voice input voice output communication aid for people with severe dysarthria. In: Proc. European Conference for the Advancement of Assistive Technology in Europe (2007)
8. English Computerized Learning Inc.: Pronunciation Power Speech Test. http://www.englishlearning.com/products/pronunciation-power-speech-test/
9. Dalby, J., Kewley-Port, D.: Explicit Pronunciation Training Using Automatic Speech Recognition Technology. Computer-Assisted Language Instruction Consortium (CALICO) Journal, vol. 16 (1999)
10. Lesson Nine GmbH: Babbel. http://es.babbel.com/#Reconocimiento-de-voz

1. Rosetta Stone: Rosetta Stone Version 4 TOTALe. http://www.rosettastone.com /content/rosettastonecom/en.html
2. Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S.: The Development and Evaluation of a Speech-to-Sign Translation System to Assist Transactions. Int. J. Hum. Comput. Interaction, 16(2):141–161 (2003)
3. Instituto Nacional de Estadística y Geografía (INEGI): Hablantes de Lengua Indígena en México. http://cuentame.inegi.org.mx/poblacion/lindigena.aspx?tema=P
4. Academia de la Lengua Mixteca: Bases para la Escritura de tu'un savi. Colección Diálogos: Pueblos Originarios de Oaxaca, México (2007)
5. Mindek, D.: Mixtecos: Pueblos Indígenas del México Contemporáneo. Comisión Nacional para el Desarrollo de los Pueblos Indígenas (2003)
6. Ferguson de Williams, J.: Gramática Popular del Mixteco del Municipio de Tezoatlán, San Andrés Yutatío, Oaxaca. Instituto Lingüístico de Verano, A.C. , México D.F. (2007)
7. Beaty de Farris, K., García, P., García, R., Ojeda, J., García, A., Santiago, A.: Diccionario Básico del Mixteco de Yosondúa, Oaxaca. Instituto Lingüístico de Verano, A.C., México, D.F. (2004)
8. Stark, S., Johnson, A., González de Guzmán, B.: Diccionario Básico del Mixteco de Xochapa, Guerrero. Instituto Lingüístico de Verano, A.C., México, D.F. (2003)
9. Instituto Lingüístico de Verano en México: Familia Mixteca. http://www.sil.org/ mexico/mixteca/00e-mixteca.htm.
10. Instituto Nacional de Lenguas Indígenas: Catálogo de las Lenguas Indígenas Nacionales: Variantes Lingüísticas de México con sus autodenominaciones y referencias geoestadísticas. http://www.inali.gob.mx/pdf/CLIN_completo.pdf (2008)
11. Anderson, L. Alejandro, R.: Vocabulario de los verbos de movimiento y de carga: Mixteco de Alacatlatzala, Guerrero. Instituto Lingüístico de Verano, A.C. http://www.sil.org/americas/mexico/mixteca/alacatlatzala/P001-Vocab-MIM.pdf (1999)
12. García, A., Miguel, R.: Nadakua'a Ndo Tee Ndo Tu'un Ndo: Aprendamos a escribir nuestro idioma. Instituto Lingüístico de Verano, A.C., México, D.F. (1998)
13. Morales, M. North, J.: Ná Cahví Tuhun Ndáhv Ta Ná Cahyí Ña: Vamos a leer y escribir en mixteco (Mixteco de Silacayoapan, Oaxaca). Instituto Lingüístico de Verano, A.C., México, D.F. (2000).
14. Alexander, R.M.: Mixteco de Atatlahuca. Instituto Lingüístico de Verano. México, D.F. (1980)
15. Pineda, L.A., Castellanos, H., Cuétara, J., Galescu, L., Juárez, J., Llisterri, J., Pérez, P., Villaseñor, L.: The Corpus DIMEx100: Transcription and Evaluation. Language Resources and Evaluation. 44: 4, 347-370 (2010)
16. Young, S., Woodland, P.: The HTK Book (for HTK Version 3.4). Cambridge University Engineering Department, Great Britain (2006)
17. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. Proc. of IEEE, 37, 257-286 (1989)
18. Jurafsky, D., Martin, J.H.: Speech and Language Processing. Pearson: Prentice Hall (2009)
19. Gillick, L., Cox, S.J.: Some statistical issues in the comparison of speech recognition algorithms. In Proc. IEEE Conf. on Acoustics, Speech and Signal Processing, 532-535 (1989)