PRONUNCIATION LEARNING SYSTEM FOR THE 32 VOWEL SYSTEM OF NASA YUWE LANGUAGE

Roberto Naranjo
University of Cauca, Colombia
rnaranjo@unicauca.edu.co

Laurent Besacier
University of Grenoble, France
laurent.besacier@imag.fr

Tulio Rojas
University of Cauca, Colombia
trojas@unicauca.edu.co

Egidio Marsico
University of Lyon 2, France
Egidio.marsico@ish-lyon.cnrs.fr

ABSTRACT

Nasa Yuwe is an indigenous language from Colombia (South America), it is, to some extent, an endangered language. Different efforts have been done to revitalize it, the most important of which being the unification of the Nasa Yuwe alphabet. The Nasa Yuwe vowel system has 32 vowels contrasting in nasalization, length, aspiration and glottalization, causing great confusion for the learner. In order to support the correct learning of this language, three classifier models (K-nearest neighbor, multilayer neural networks and Hidden Markov Model) have been developed to detect confusion in the pronunciation of the 32 vowels. They were developed in three different experiments in order to reach the best accuracy rates. The selected strategy developed binary classifiers using bagging with adding a number of negatives samples for each vowel, with an accuracy rate of about 85%. With these trained classifiers, a Computer Assisted Language Learning system prototype (**CALL**) was designed to support the correct pronunciation of the language's vowels. Additionally using this system, the native and non-native speakers score distribution of acceptance was calculated and the confusion of vowels for non-native speaker corpus was evaluated.

Index terms- Vowel classification, Nasa yuwe, pronunciation learning, pattern recognition, computer assisted language learning.

1. INTRODUCTION

The indigenous Nasa people are the second largest ethnic group in Colombia (South America) with predominance in the Department of Cauca. The knowledge of the Nasa people is vehiculed through oral tradition, mainly by the elders of the community (Pebi, 2000). The Nasa community is now gradually losing the use of their language, even though efforts have been made to maintain their oral tradition over time. This work is part of a broader project developing an online virtual Nasa community in order to allow Nasa language and people to adapt to modern life (Sierra et al., 2010). It includes a module of computer assisted language learning based on the acoustic model presented here. The Nasa Yuwe vowel system is organized around four vowel qualities (i, e, a, u), with a primary opposition between oral and nasal vowels. These two series can then contrast in length, glottalization and aspiration making a total of 32 distinctive vowels, 16 oral and 16 nasal (Marsico & Rojas Curieux, 1998, Rojas Curieux 1998). The many acoustic distinctions are a challenge for the learner both in production and perception especially for cues like nasality or glottalization which can vary a lot in their realizations. In order to build a Computer Assisted Language Learning software prototype (**CALL**) based on pattern recognition techniques, three different experiments were conducted in order to detect the vowels confusion. The first strategy developed a classifier for each vowel group, the second one developed a binary classifier for each vowel and the third one developed a binary classifier for each vowel using bagging with adding a number of negative samples. This last strategy was selected because it showed the best accuracy rates. The K-nearest neighbor, multilayer neural networks and Hidden Markov Model classifiers were taken into account for this task. Therefore, this paper presents classifiers for the 32 vowels in order to design a CALL for the correct pronunciation of the vowels of Nasa Yuwe. Section 2 of this paper presents the methodology, section 3 the background, section 4 the modeling of vowels, section 5 the description of the system prototype and its evaluation, and section 6 the conclusions and future work.

2. METHODOLOGY

2.1. Description of the vowels

There are four Nasa yuwe vowel group: A, E, I, and U. Each group is subdivided into oral and nasal, and within each division there are four modes of articulation: simple, glottalized, aspirated and lengthened, therefore producing 32 classes (in 4 vowel groups, each group with 8 members). Table 1 (Rojas, 2001) shows the Nasa yuwe vowels represented by IPA symbols (IPA, 2011).

Table 1. Nasa yuwe vowels

	Oral Vowel				**Nasal Vowel**			
Simple	a	e	i	u	ã	ẽ	ĩ	ũ
Glottal	a'	e'	i'	u'	ã'	ẽ’	ĩ'	ũ'
Aspirated	a^h	e^h	i^h	u^h	$ã^h$	$ẽ^h$	$ĩ^h$	$ũ^h$
Elongated	a:	e:	i:	u:	ã:	ẽ:	ĩ:	ũ:

2.2. Steps

To do this work, the following steps were conducted: as a first step, a collection of spoken and written texts of the language were gathered. As a second step were calculated 30 linear prediction coefficients (LPC) (Huang et al., 2001) and the residual energy for each vowel segment. Then, one have to choose a classifier model, this is done after evaluating three classifiers, within which the alternatives considered are K-nearest neighbor classifier (KNN) (Kuncheva, 2004), multilayer neural networks classifier (MLP) (Haykin, 1998), and the Hidden Markov Models (HMM) (Rabiner et al., 1993).

3. BACKGROUND IN CALL

A CALL (Computer Assisted Language Learning) system helps a human learner to improve its language skills, and supports teaching and learning of a second language. Such a system has several advantages for both teachers and students, like the possibility to practice the material many times in an environment free of stress (Wang et al., 2009). For Nasa yuwe there is no history of building such systems, therefore previous works use models that have been tested for other languages for detecting pronunciation errors. In Franco et al., (1999) are presented models based on calculating the likelihood ratio test (LRT) that uses a database of labeled speakers to produce two acoustic models for each phoneme. The first model called λc is produced with correct pronunciation (native speakers), while the second model called λm is done with incorrect pronunciation (non-native speakers). Then for each segment *q* belonging to a pronunciation *s* and *o* the observation, the Likelihood Ratio denoted LLR(o, q) is calculated using the correct and incorrect acoustic models. Finally, the LLR(o, q) score is compared with a phoneme-dependent threshold to detect if the segment *q* is pronounced correctly or not. In Witt et al., (2000) a model of goodness of pronunciation (GOP) is proposed to detect errors in pronunciation, which is a variation on the model of posterior probability (Franco et al., 1999). According to the proposal, the mean and the variance obtained by the data analysis can be used to select an appropriate threshold for each phoneme. A pronunciation below this threshold is considered as good while a higher one is tagged as an error of pronunciation. In Troun et al., (2009) acoustic-phonetic models with binary classifiers are used, comparing two pattern recognition techniques such as decision trees and linear discriminant analysis, to classify the sounds that cause more problems with pronunciation, fricative / x / and occlusive / k / in Dutch. The first method proposed consists of a decision tree classifier and the feature vector is formed by peak Race of Rice (ROR= $(E_n - E_{n-1})$ $/\Delta t$, for each window n the amplitude E_n is measured by computing the logarithm of the Root-Mean-Square over window n). If the peak ROR is above a threshold, it is considered occlusive, otherwise it is considered fricative. This method achieved an accuracy rate between 75% and 91%. The second method used the linear discriminant classifier and the feature vector is formed by amplitude characteristics, higher ROR, and wave duration, this method achieved an accuracy rate of between 85% and 95% of classification.

4. VOWEL MODELING FOR NASA YUWE

The following describes the task of detecting confusion in the pronunciation of vowels in Nasa yuwe language, exploring various classification methods in order to find the most accurate classifiers for this task.

4.1. Corpus Construction

A group of 250 words is selected, which have a phonetic structure made up of vowel-consonant (VC), consonant-vowel (CV), consonant-vowel-consonant (CVC), and consonant-consonant-vowel-consonant-vowel (CCVCV). From these words, 4.224 recordings were collected with five native speakers, three men and two women, and for each one of the 32 vowels, 132 repetitions were obtained, setting up the native speaker corpus. The corpus is sampled at 44.1 kHz in mono format and processed from 0db to 60db. In the same way, 1088 recordings were obtained from three non-native speakers (two men and one woman, they have low skills level of Nasa Yuwe language) setting up the non-native corpus.

4.2. Features extraction

A Centroid for each word occurrences was obtained on native speaker corpus using Dynamic Time Warping (DTW) (Sakoe et al., 1978) (Dtw Matlab, 2011). All aligned paths were averaged, giving a centroid for each word (Casacuberta et al., 1991). Then the start and end of the vowel present in each of the centroids were found using a reference word (selected randomly among all the word occurrences). With DTW the reference word and the centroid are aligned. The advantage of having a centroid for each word is that it helps to segment the vowels of the whole corpus automatically using DTW between the centroid and every other word utterance. Such DTW procedure could also be used in the CALL prototype. The feature vectors were formed by 30 LPC coefficients and the residual energy after down-sampling to 16 kHz each segment of each vowel.

4.3. Classification Experiments

As noted in Section 2, an analysis of various classification methods is involved. The idea is to use the same information with all of them to establish experimentally which one obtains the best accuracy rate ({True Positives + True Negatives} / {True Positives + True Negatives + False Positives + False Negatives}) (Kuncheva, 2004) in

characterizing the vowels of Nasa yuwe. The MathLab 7.0 toolbox was used (Matlab, 2011) as a platform for experimentation, using specifically the libraries PrTools (Prtools, 2011) and HMM toolbox (Ghahramani, 2011). The project conducted three experiments; the results obtained are described below.

4.3.1. First Experiment: a Classifier for each vowel group (a, e, i, u)

The first step in constructing a classifier model is to develop a classifier separately for each of the four groups of vowels (a, e, i, u). Input into the classifier consists of LPC coefficients of each sample and there are 8 outputs that correspond to each class of the same vowel group. Three classifiers were trained and tested such as: KNN (with one nearest neighbors, 1-NN), HMM (three states, interconnected, with 10 cycles of Baum-Welch) and MLP (an input layer, two hidden layers each with 25 neurons, an output layer, training algorithm was Levenberg-Marquardt (Levenberg, 1944) (Marquardt, 1963), and a training epoch of 50). The data sets for this experiment are as follows: 2 data sets created by vowel group, the first data set is for training data with 80% of randomly selected samples (106 objects per vowel) and the second one is for testing data with 20% (26 objects per vowel), the classifiers were trained and tested by 20-folds cross-validation.

Tables 2, 3, 4, 5, show the average accuracy rate obtained per vowel and classifier. As a result of this experiment, we found that 1-NN is the best classifier for each vowel group, with the following accuracy rates: for the A vowel group, 50.48%, for the E vowel group, 63.82%, for the I vowel group, 63.77%, and for the U vowel group, 58.46%. Also, the accuracy values obtained for the HMM classifier in the A, E and U vowel groups are quite close a 1-NN. In this experiment the average accuracy rate obtained is between 50% and 63%, hence an individual classifier is not suitable to classify the 8 vowels in the same vowel group because the accuracy rates obtained are low, for this reason it's necessary to define another training strategy to improve the average accuracy rates obtained, therefore we carried out a second experiment.

4.3.2. Second Experiment: a Binary classifier for each of the 32 vowels

In this experiment binary classifiers are trained for each of the 32 vowels. The same classification task as in the first experiment is explored and the best for each vowel is selected. Each classifier has as inputs the LPC coefficients and an output that indicates whether the entry belongs to the class or not. The data for this experiment are comprised of 32 data sets, the training data, with each training set consisting of 106 samples of each vowel representing the voice of the positive samples and labeled with their respective class between 1 and 32, and also contains 106 randomly-selected objects from every other vowels of the same vowel group, representing the negative vowel samples and labeled as class 33. The following sets correspond to the test data, 26 positive samples labeled with the respective class between 1 and 32 were taken of each vowel, and also for the negative samples the same number of vowel samples were taken and labeled with class 33. The classifiers were ıined and tested by 20-folds cross-validation, the following the analysis of results.

Table 2. Results of experiment 1, with individual classifiers in the A vowels group

VOWEL	KNN	MLP	HMM
a	23,85%	32,79%	21,15%
a'	21,54%	27,94%	40,96%
ã	44,23%	39,88%	14,23%
ã'	95,19%	58,30%	11,15%
aʰ	70,19%	50,81%	72,88%
ãʰ	97,12%	5 ,83 %	77,31%
a:	30,96%	27,53%	56,54%
ã:	20,77%	23,08%	88,46%
Average	**50,48 %**	**39,14 %**	**47,84 %**

Table 3. Results of experiment 1, with individual classifiers in the E vowels group

VOWEL	KNN	MLP	HMM
e	35,38%	37,69%	11,15%
e'	97,12%	64,42%	39,04%
ẽ	35,77%	31,73%	17,12%
ẽ'	25,19%	21,92%	9,81%
eʰ	95,96%	59,04%	7,31%
ẽʰ	70,19%	67,31%	92,12%
e:	100,0%	60,77%	71,15%
ẽ:	63,82%	34,62%	100,0%
Average	**63,82 %**	**47,19 %**	**43,46 %**

Table 4. Results of experiment 1, with individual classifiers in the I vowels group

VOWEL	KNN	MLP	HMM
i	33,27%	28,08%	51,54%
i'	52,50%	55,77%	63,08%
Ĩ	79,42%	49,23%	50,19%
Ĩ'	97,88%	75,77%	36,73%
iʰ	67,50%	64,62%	24,23%
ĩʰ	31,73%	31,35%	85,00%
i:	95,96	79,62%	68,46%
ĩ:	51,92%	48,27%	96,54%
Average	**63,77 %**	**54,09 %**	**59,47 %**

Table 5. Results of experiment 1, with individual classifiers in the U vowels group

VOWEL	KNN	MLP	HMM
u	26,35%	45,96%	16,73%
u'	67,50%	50,00%	59,62%
ũ	30,00%	32,12%	27,50%
ũ'	95,19%	45,19%	30,77%
uʰ	70,00%	48,85%	35,77%
ũʰ	98,27%	43,08%	90,58%
u:	24,1%	19,04%	75,96%
ũ:	55,58%	24,23%	92,88%
Average	**58,46 %**	**38,56 %**	**53,73 %**

For the A vowel group (see Table 6), the MLP and HMM classifiers are the best option because they present the best accuracy rate for each vowel and this has a distribution between 62.31% and 93.56%. For E vowel group (see Table 7), the MLP classifier reports the best accuracy rate for most vowels and the accuracy rate distribution is between 66.73% and 100%, only for ẽ: vowel the HMM classifier is better. HMM classifier is the best for i', ĩʰ, i: and ĩ: vowels (see Table 8), for iʰ vowel, KNN classifier is the best, for every others, the MLP classifier is better. The accuracy rate distribution is between 70.29% and 96.06%. Finally, for U vowel group, MLP and HMM classifiers are the best option (see Table 9), the accuracy rate is distributed between

65.10% and 98.85%. In this experiment, we obtained accuracy rates between 62.31% and 100%, and there is also a variety of classifiers that show the best accuracy rate for each vowel.

Table 6. Results of experiment 2, with individual classifiers in the A vowels group

VOWEL	KNN	MLP	HMM
a	66,15%	70,96%	65,19%
a'	66,35%	69,81%	70,77%
ã	53,17%	62,31%	49,23%
ã'	62,98%	63,08%	43,27%
aʰ	67,60%	72,50%	75,77%
ãʰ	86,54%	90,48%	90,67%
a:	74,90%	80,87%	82,79%
ã:	93,56%	87,69%	87,02%
Average	**71,41 %**	**74,71 %**	**70,59 %**

Table 7. Results of experiment 2, with individual classifiers in the E vowels group

VOWEL	KNN	MLP	HMM
e	61,92%	6,73%	54,2%
e'	76,25%	77,02%	83,08%
ẽ	59,90%	69,33%	65,38%
ẽ'	90,67%	92,69%	60,96%
eʰ	68,17%	73,46%	56,44%
ẽʰ	82,40%	86,06%	83,85%
e:	80,38%	82,31%	82,12%
ẽ:	94,23%	93,75%	100,0%
Average	**76,74 %**	**80,17 %**	**73,29 %**

Table 8. Results of experiment 2, with individual classifiers in the I vowels group

VOWEL	KNN	MLP	HMM
i	63,65%	70,29%	64,33%
i'	78,27%	84,23%	84,71%
Ĩ	62,88%	70,10%	67,31%
Ĩ'	73,75%	85,00%	81,83%
iʰ	79,81%	79,33%	55,38%
ĩʰ	90,38%	91,63%	96,06%
i:	7,75%	79,90%	80,48%
ĩ:	88,56%	0,58%	92,60%
Average	**76,38 %**	**81,38 %**	**77,84 %**

Table 9. Results of experiment 2, with individual classifiers in the U vowels group

VOWEL	KNN	MLP	HMM
u	63,08%	71,54%	54,23%
u'	80,48%	76,92%	80,58%
ũ	61,25%	70,58%	61,54%
ũ'	70,96%	82,60%	72,02%
uʰ	61,54	65,10%	60,58%
ũʰ	87,60%	89,62%	92,50%
u:	81,54%	84,62%	71,92%
ũ:	91,06%	91,54%	98,85%
Average	**74,69 %**	**79,06 %**	**75,28 %**

4.3.3. Third Experiment: Use of bagging and adding the number of negative samples

In this experiment individual and bagging KNN, MLP and HMM classifiers were trained and tested by 20 folds cross-validation for each vowel, increasing the negative samples to 742 for each vowel, because this allows the classifier to improve its ability to reject negative samples and therefore increasing the accuracy rate; on the other hand, we had no more positive samples. The scheme bagging (bagging selects randomly n-subsets from data set, then with each subsets, n-classifiers are trained, and finally the n-classifiers are combined by voting) (Kuncheva, 2004) is appropriate because it applies to low-performing classifiers, unstable classifiers and for small corpora. The following is the analysis of results.

Tables 10, 11, 12, 13, present the better results for Bagging or Individual classifiers, for example in table 10, individual KNN, Bagging MLP and Bagging HMM classifiers got the best accuracy rates for A vowels group.

For A vowel group in some vowels the Bagging MLP classifier got the best accuracy rate (see Table 10), in others cases the KNN classifier or Bagging HMM classifier were best, for example the ã: vowel got an accuracy rate of 98.08% with the KNN classifier, for this vowels group the accuracy rate distribution is between 62.12% and 98.08%. For some E vowels, the Bagging KNN classifier got the best accuracy rate (see Table 11), in other cases the MLP or Bagging HMM classifiers were better, for this vowels group the accuracy rate distribution is between 65.38% and 100%. For most of I vowels (see Table 12), the Bagging MLP classifier got the best average accuracy rate, but for ĩ and ĩʰ vowels, the Bagging HMM and Bagging KNN classifiers were better respectively. For this vowel group the accuracy rate distribution is between 76.92% and 100%. Finally, for U vowel group (see Table 13), the accuracy rate distribution is between 69.33% and 100%.

Table 10. Results of experiment 3, in the A vowels group

VOWEL	KNN	BAG MLP	BAG HMM
a	53,85%	69,23%	67,40%
a'	57,69%	65,38%	76,73%
ã	55,77%	55,77%	62,12%
ã'	57,69%	67,31%	70,87%
aʰ	65,38%	84,62%	72,60%
ãʰ	92,31%	96,15%	81,35%
a:	73,08%	76,92%	73,27%
ã:	98,08%	94,23%	85,67%
Average	**69,23 %**	**76,20 %**	**73,75 %**

Table 11.Results of experiment 3, the E vowels group

VOWEL	BAG KNN	MLP	BAGG HMM
e	65,38%	63,46%	62,31%
e'	86,54%	82,69%	71,25%
ẽ	65,38%	63,46%	61,92%
ẽ'	98,08%	98,08%	90,29%
eʰ	67,31%	75,00%	59,13%
ẽʰ	90,38%	92,31%	85,29%
e:	78,85%	84,62%	91,06%
ẽ:	96,15%	100,0%	96,63%
Average	**81,01 %**	**82,45 %**	**77,24 %**

Table 12. Results of experiment 3, in the I vowels group

VOWEL	BAG KNN	BAG MLP	BAG HMM
i	67,31%	76,92%	69,52%
i'	71,15%	82,69%	82,02%
ĩ	71,15%	71,15%	77,79%
ĩ'	71,15%	86,54%	79,71%
iʰ	88,46%	92,31%	82,50%
ĩʰ	100,0%	96,15%	96,54%
i:	8,46%	88,46%	83,46%
ĩ:	98,08%	98,08%	88,65%
Average	**81,97 %**	**86,54 %**	**82,52 %**

Table 13. Results of experiment 3, in the U vowels group

VOWEL	BAG KNN	MLP	BAG HMM
u	75,00%	67,31%	64,23%
u'	86,54%	8,85%	80,19%
ũ	65,38%	69,23%	69,33%
ũ'	71,15%	88,46%	5,67%
uʰ	65,38%	67,31%	78,85%
ũʰ	98,08%	100,0%	85,00%
u:	82,69%	82,69%	77,60%
ũ:	96,15%	100,0%	87,60%
Average	**80,05 %**	**81,73 %**	**77,31 %**

The average accuracy rates obtained in this experiment are better than the second experiment, therefore these classifiers have been selected to build the software prototype; table 14 shows the chosen classifiers list for each vowel. Note that the bagging schema worked for most of the vowels and especially with MLP and HMM, additionally, we found for A vowel group an average accuracy rate of 79.34%, for E vowel group an average accuracy rate of 84.22%, and for I and U vowel groups an average accuracy rate of 87.85% and 85.11%, respectively.

Table 14. Chosen classifiers list for each vowel

Vowel	Oral				Nasal			
	Simple	glottal	Aspirated	Elongated	Simple	glottal	Aspirated	Elongated
A	Bagging MLP	Bagging HMM	Bagging MLP	Bagging MLP	Bagging HMM	Bagging HMM	Bagging MLP	KNN
E	Bagging KNN	Bagging KNN	MLP	Bagging HMM	Bagging KNN	Bagging KNN	MLP	MLP
I	Bagging MLP	Bagging MLP	Bagging MLP	Bagging MLP	Bagging HMM	Bagging MLP	Bagging KNN	Bagging MLP
U	Bagging KNN	Bagging KNN	Bagging HMM	Bagging MLP	Bagging HMM	Bagging MLP	Bagging MLP	Bagging MLP

5. SYSTEM PROTOTYPE AND ITS EVALUATION

A support software prototype was built for the correct pronunciation of Nasa yuwe vowels, enabling a human learner to practice the pronunciation of these vowels. The software prototype has a menu with the four vowel groups. Once the vowel group is chosen, it shows the 8 vowels of that group. Once the user has selected a vowel, they are presented with a menu where the words associated with the chosen vowel are found and the user can select one. After the user chooses a word, the software plays the recording of the word found in the corpus and the user immediately does his pronunciation. The system then extracts the vowel present in the spoken word by DTW into segments using the centroid of the word and this segment is analyzed using a set of previously developed classifiers (see table 14), determining whether it is the correct or wrong pronunciation. In the case of incorrect pronunciation, the two vowels more likely to be confused are determined, in both cases there is a message to the user. An evaluation of the system was made on non-native and native speaker corpus, bellow the results are presented.

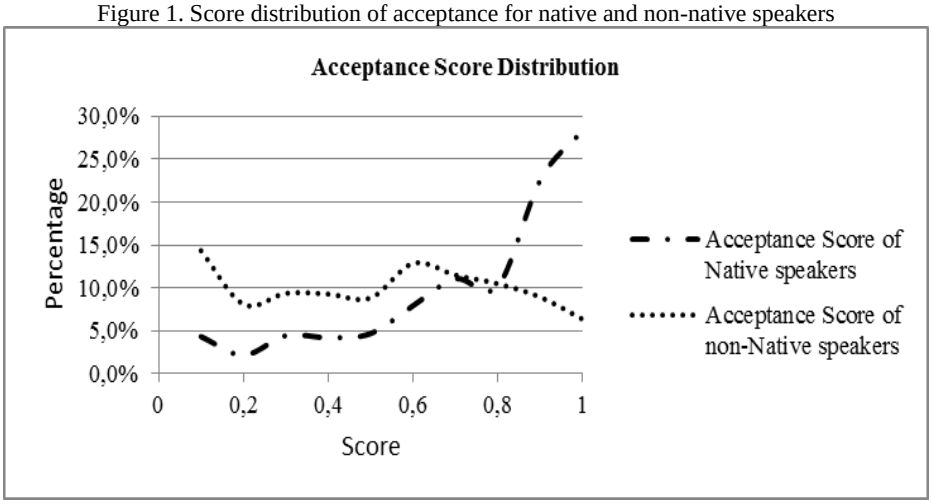
5.1. Score distribution of acceptance

Figure 1 shows the score distribution of acceptance that was calculated on the native (five native speakers with 845 utterances) and non-native (three non-native speakers with 1088 utterances) speaker corpus by the system. Noted that the distribution between 0.0 and 0.7 corresponds to the low scores, the non-native speakers distribution is higher than native speakers distribution, this means that a high percentage of non-native speaker pronunciations have low scores. The opposite situation happens with the higher scores (above 0.7) where the native speaker distribution is higher than non-native speaker distribution. According to the score distribution, many native speaker pronunciations have scores higher than non-native speaker pronunciations, therefore we see how the system assigned two different score distributions of acceptance to native and non-native speakers. In the future, this should allow us to establish pronunciation levels of non-native speakers according to the native speakers.

5.2. Analysis of the confusions

In this section we present a brief analysis of some confusion made by the system on the non-native speakers corpora, in order to determine whether they were mistakes of the system or of the learners. We focused on the mistakes regarding the a' vowel. Out of the 38 utterances analyzed by the system 9 were correctly classified, 9 were classified as a, 9 were classified as ã, 6 as ã', 3 as a: and 2 utterances were classified as ã:. For the 9 utterances classified by the system as a, 7 were pronounced by the human learner as a, therefore the system has correctly identified the human learner's mispronunciation and the other two pronunciations were pronounced by the human learner as a' (system errors). However, we have checked two native speaker pronunciations of the corresponding a' vowel and we have found an important degree of variation in the realization of the glottalization. This can undoubtedly affect the acoustic model of the system and thus the correct identification of these vowels. For the 9 utterances classified by the system as ã, 2 utterances were pronounced by the human learner as ã (human learner's mispronunciation), 4 utterances were pronounced as a, and 3 utterances were pronounced as a', these 7 are considered as system errors. But again, we have checked the native speaker corpus and identified a low level of nasality in the ã' vowel of one native speaker woman, this variation in the native speaker corpus could explain the system errors. In conclusion, we believe our system should help CALL for these particulars vowels, however it seems that a detailed acoustic characterization of the language would be helpful.

Figure 1. Score distribution of acceptance for native and non-native speakers



6. CONCLUSIONS AND FUTURE WORK

Although Nasa Yuwe vocalic system seems hard to capture due to the subtle acoustic differences between its 32 vowels, our model manage to reach promising accuracy rates. This is achieved by making use of a binary classifier using bagging and adding the number of negative samples for each vowel. The accuracy rates on average that the system have are around 85%, meaning that the rates are well above this value as in the case of vowels: ãʰ, ãː, ẽˈ, ẽː, ĩʰ, ĩː, ũʰ and ũː, whose accuracy rates are above 95%, with the opposite occurring with the vowels ã, e, ẽ and ũ which are below 70%. For all other vowels, accuracy rate is between 70% and 95%, with these rates being within those found in other related projects (see Section 3). Multilayer neural networks and Hidden Markov Model are for most vowels the best classifier, in the case of MLP, the most appropriate configuration is of two hidden layers with 25 neurons in each hidden layer and in the case of HMM, 3-states were appropriated. The score distribution of acceptance was calculated by the system, which shows a high percentage of non-native speaker pronunciations with low acceptance scores, opposite situation occurs with the native speaker pronunciations, a high percentage have high acceptance scores, in the future, this should allow us to establish pronunciation levels. Finally, we found that there is variation in the native speaker corpus and this explains the system errors when it classifies the non-native speaker corpus.

For future work, we suggest using other training strategies for the classifiers such as AdaBoost (Freund, 1997), increasing the native speaker corpus at least four times, to carry out a detailed acoustic characterization of the language, and using special equipment like a nasograph to better capture the features of nasality.

7. REFERENCES

[1] Casacuberta, F., Vidal, E., Aibar, P., Decodificación Acústico Fonética mediante plantillas subléxicas. Procesamiento del lenguaje natural, Nº. 11, 1991, pp. 265-274.

[2] Duda, O., Hard, R., Stork, P., Pattern Classification, 2 Ed. Jhon Wiley & Son, 2001.

[3] Dtw Matlab, http://labrosa.ee.columbia.edu/matlab/dtw/, Last access July 2011.

[4] Franco, H., Neumeyer, L., Kim, Y., Ronen, O., Bratt, H., Automatic detection of phone-level mispronunciation for language learning, In: Proc. European Conference on Speech Communication and Technology, 1999, pp. 851–854.

[5] Freund, Y., Schapire, R., A decision theoretic generalization of On line learning and an application to Boosting, Journal of computer and system sciences 55, 1997, pp. 119-139.

[6] Ghahramani, Z. http://www.gatsby.ucl.ac.uk/~zoubin/software.html, last access November 2011.

[7] Haykin, S., Neural Networks: A Comprehensive Foundation (2nd Edition), Prentice Hall, 1998.

[8] Huang, X., Acero, A., Hon, H., Spoken Language Processing, Ed. Prentice Hall, 2001, pp. 290-303.

[9] International Phonetic Alphabet (IPA). http://ipa.typeit.org/. Last access July 2011.

[10] Kuncheva, L., Combining pattern classifiers: models and algorithms, Ed. Jhon Wiley, 2004.

[11] Levenberg, K. A Method for the Solution of Certain Problems in Least Squares. Quart. Appl. Math. **2,** 1944, pp. 164-168.

[12] Marsico, E., Rojas, T., Etude acoustique préliminarire des 16 voyelles orales du Paez de Talaga, langue amérindienne, XXII journees d'etude sur la parole, 1998.

[13] Marquardt, D., An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM J. Appl. Math. 11, 1963, pp. 431-441.
[14] Matlab, http://www.mathworks.cm/products/matlab/, last access July 2011.
[15] Pebi (Programa de educación Bilingüe - Cric)., Acerca de la unificación del alfabeto Nasa yuwe, Revista C'ayu'ce, No 4, 2000, pp. 52-53.
[16] Prtools, http://www.prtools.org/, last access July 2011.
[17] Rabiner, L., Juang, L.B., Fundamental Speech Recognition, Prentice - Hall International Inc, 1993.
[18] Rojas, T., Desde arriba y por abajo construyendo el alfabeto nasa. La experiencia de la unificación del alfabeto de la lengua Páez (nasa yuwe) en el Departamento del Cauca – Colombia. http://lanic.utexas.edu/project/etext/llilas/cilla/rojas.html. 2001, last access July 2011.
[19] Rusell, S., Norving, P., Inteligencia Artificial- un Enfoque Moderno. 2 Ed. Editorial Prentice Hall, 2004.
[20] Sakoe, H., Chiba, S., Dynamic programming optimization for spoken word recognition, IEEE Trans, Acoust, Speech Signal Process, Vol. ASSP-26, No. 1, 1978, pp. 43-49.
[21] Sierra, L., Naranjo, R., Rojas, T., Ewa: Comunidad virtual de apoyo a procesos de Etnoeducacion Nasa. Ed. Universidad del Cauca, 2010.
[22] Troun, K., Neri, A., Cuacchiarini, C., Strik, H., Automatic pronunciation error detection: an acoustic-phonetic approach, University of Nijmegen. http://citeseerx.ist.psu.edu/, 2009, last access July 2011.
[23] Wang, H., Christopher, J., Waple, J., Kawahara, T., Computer Assisted Language Learning system based on dynamic question generation and error prediction for automatic speech recognition, Science Direct, Speech Communication 51, 2009, pp. 995–1005.
[24] Witt, S.M., Young, S.J., Phone-level pronunciation scoring and assessment for interactive language learning, Speech Comm 30, 2000, pp. 95–108.