Developing Resources for Automated Speech Processing of the African Language Naija (Nigerian Pidgin)

Brigitte Bigi, Bernard Caron, Oyelere S Abiola

Developing Resources for Automated Speech Processing of the African Language Naija (Nigerian Pidgin)

Brigitte Bigi*, **Bernard Caron**†**, **Oyelere S. Abiola****

*LPL, CNRS, Aix-Marseille Univ. brigitte.bigi@univ-amu.fr
†IFRA-Nigeria, USR3336, CNRS. bernard.caron@cnrs.fr
** University of Ibadan. biolaoye2@gmail.com

Abstract

The development of HLT tools inevitably involves the need for language resources. However, only a handful number of languages possesses such resources. This paper presents the development of HLT tools for the African language Naija (Nigerian Pidgin), spoken in Nigeria. Particularly, this paper is focusing on developing language resources for a tokenizer, an automatic speech system for predicting the pronunciation of the words and their segmentation.
The newly created resources are integrated into SPPAS software tool and distributed under the terms of public licenses.

1. Introduction

The development of Human Language Technologies (HLT) tools is a way to break down language barriers. There are approximately 6000 languages in the world but unfortunately, only a handful possess the linguistic resources required for implementing HLT technologies (Bigi, 2014). Large corpora datasets for most of the under-resourced languages are created by HLT researchers for Natural Language Processing (NLP) or for Speech Technologies. African languages form about 30% of the world languages and their native speakers form 13% of the world population (Lewis et al., 2015). However, "NLP in Africa is still in its infancy; of about 2000 languages, a very few have featured in NLP research and resources, which are not easily found online." (Onyenwe, 2017). With a 182M population, Nigeria counts about 92M Internet users for June, 2015, i.e. 51.1% of its population[1] and it's constantly growing. The official language is English but over 527 individual languages are spoken in Nigeria[2]. Recently, the Igbo language was one of the Nigerian languages investigated for NLP (Onyenwe, 2017).

Among these Nigerian languages, Nigerian Pidgin English (NPE) is a post-creole continuum that is spoken as a first language (L1) by 5 million people, while over 70 million people use it as a second language (L2) or as an inter-ethnic means of communication in Nigeria and in Nigerian Diaspora communities. The origin of NPE is generally described as a development out of an English-lexified jargon attested in the 18th Century in the coastal area of the Niger delta (River State), with some lexical influence from Krio through the activities of missionaries from Sierra Leone (Faraclas, 1989). The heartland of NPE is the Niger Delta, with Lagos and Calabar as secondary extensions. NPE is identified by "pcm" in the iso-639-3 language codes.

Since the independence of Nigeria in 1960, NPE has been rapidly expanding from its original niche in the Niger delta area, to cover two-thirds of the country, up to Kaduna and Jos, and is now deeply rooted in the vast Lagos conurbation of over 20 million people. It has become, over the last 30 years the most important, most widely spread, and perhaps the most ethnically neutral lingua franca used in the country today. In this geographical expansion, and as it conquers new functions (e.g. in business, on higher education campuses, in the media and in popular arts), NPE is subject to extensive contact and influence from its original lexifier, i.e. English and from the multitude of vernacular Nigerian languages. A mixed language has emerged that is fast expanding (both in geography and function) and rapidly changing. The name **Naija** (meaning 'Nigeria' in NPE) is used to describe this language learnt and used as an L2 in most of Nigeria, and differenciate it from the creolised variety (NPE) spoken as an L1 in the Niger delta (see (Esizimetor and Egbokhare, 2012) for a short characterization of Naija). Naija is the object of this paper on the development of HLT as part of the NaijaSynCor project[3]. It aims at describing the language in its geographical and sociological variations, based on a 500k word corpus annotated and analyzed with cutting edge HLP tools developed for corpus analysis.

Then, the development of speech technologies for Naija faces the following problems: (1) lack of language resources (lexicon, corpora, ...), not to mention digital resources; and (2) acoustic and phonological characteristics that still need to be properly investigated. These issues are shared with most under-resourced languages, and linguists are currently looking for solutions to solve or to avoid them. Nevertheless, language data collection is still a challenging and fastidious task.

This paper describes the development of a corpus and some language resources for Naija as part of a corpus-based project, which aims at evaluating the nativization of the language. Such newly created linguistic resources were integrated into SPPAS software tool (Bigi, 2015) for a tokenizer, an automatic speech system for predicting the pronunciation of words and their segmentation.

[1] source: http://www.internetworldstats.com/af/ng.htm - 2017-10

[2] source: https://www.ethnologue.com/country/NG - 2017-06

[3] http://naijasyncor.huma-num.fr/

2. Corpus creation

For the NaijaSynCor project, a total of 384 samples of oral corpus (monologues, dialogues), an average 6 min each, is to be collected from 380 speakers so as to represent the widest scope of functions and locations of Naija in the country. The speech recordings are done using professional digital recorders and wireless microphones - one per speaker.

At the initial stage of the project, 8 of these recordings were partially manually transcribed and time-aligned at the phonetic level (Table 1). The transcriptions use the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) code, a machine-readable phonetic alphabet that was originally developed by (Wells, 1997). The recordings are totalling 3 min 29 seconds in length, 4 men (M) and 4 women (W). Only these files are currently available to construct HLT tools: recordings are being collected and their orthographic transcription is still in progress.

File (wav)	Recording duration (in sec.)	Speech duration (in sec.)	Nb of phon-	Speech rate (phon/sec)
M_1	32.578	20.817	254	12.20
M_2	35.155	23.509	281	11.95
M_3	48.431	35.960	403	11.20
M_4	40.243	20.213	233	11.53
W_1	33.698	28.708	360	12.54
W_2	35.926	28.174	258	9.16
W_3	37.311	27.790	284	10.22
W_4	34.239	24.087	263	10.92

Table 1: Description of the transcribed corpus, manually time-aligned at the phoneme level.

Orthographic transcription is often the minimum requirement for a speech corpus, as it is the entry point for most HLT tools. Corpora are using the orthography developed by (Deuber, 2005) in her work on Lagos Nigerian Pidgin. This etymological orthography - adapted from the lexifier language orthography, i.e. English - has been chosen preferably to the phonological script used by linguists as it is spontaneously used by educated Nigerians, and thus easier to teach to transcribers. Code-switched to English sections are identified by dedicated boundaries.

One of the characteristics of speech is the important gap between the phonological form of a word and its phonetic realizations. These specific phenomena have a direct influence on the automatic phonetization procedure as shown in (Bigi et al., 2012). The orthographic transcription convention that was adopted to transcribe Naija is the one proposed for the SPPAS software tool. Two transcriptions can then be automatically derived by the tokenizer: the "standard transcription" that is suitable for a POS-Tagger and the "faked transcription" that is a specific transcription from which the obtained phonetic tokens are relevant for a phonetization system (Bigi, 2014; Bigi, 2016). The following are examples of transcribed speech:

(W_2) '*so, all Edo people wey don travel go different-different-different places, everybody go come travel come back.*' So, all the Edo people who have travelled far and wide, everybody will return home.

(M_2) '*So, we don carry di matter come again, as we dey always carry am come.*' So, we have brought the topic again, as we always bring it.

3. Phonetic description of Naija

The phonetic transcription of a few minutes of speech enabled us to establish the list of the phonemes that are mostly used by the speakers. While the list of consonants is pretty close to the English one, the list of vowels used in Naija language is very different. As shown in Table 2, only the sounds /dZ/ and /tS/ were observed infrequently, but the English /4/, /Z/ and /D/ were not observed in the transcription. /D/ appears in the larger corpus as a variant of /d/, and could be a socio-linguistic marker of educated speakers; /Z/ is extremely rare, and seems idiosyncratic to some speakers. /T/ was also observed in the under-development larger corpus. The other consonants of English are shared by both languages. But as shown in Table 3, the set of vowels Naija and English are sharing is only: /E/, /i/, /u/ and the two diphthongs /aI/ and /aU/. Six different nasalized vowels were observed in the corpus, but with a small number of occurrences (total is 81 occurrences).

b	d	g	k	p	t	tS	dZ	Z
37	163	52	90	54	119	3	3	0
m	n	N	j	w	l	r\		4
98	140	4	26	89	57	58		0
S	f	s	z	v	h		T	D
23	48	141	12	24	12		0	0

Table 2: Inventory of consonants and their occurrences in the manually time-aligned files

a	e	E	i	o	O	u	aI	aU
203	123	111	221	93	126	74	37	8
a∼	e∼	E∼	i∼	O∼	u∼			
21	1	18	18	20	3			

Table 3: Inventory of vowels and their occurrences in the manually time-aligned files. Second line refers to nasals.

4. Creating resources for HLT tools

4.1. Pronunciation dictionary and lexicon

A pronunciation dictionary and a lexicon were manually created, including Naija specific words only. The dictionary was originally created by extracting the lexicon of the corpus published in annex of (Deuber, 2005). The observed pronunciations of the corpus of Table 1 were added to the dictionary. Gradually, as more corpora were transcribed, new words with their orthographic variants were added to the dictionary and lexicon. A team of four transcribers, native speakers of the language, were in charge of establishing the pronunciation of the words, with their most frequent variants, and transcribing them in X-SAMPA. The same team was in charge of the 8 samples (section 2.) transcribed and aligned to create the acoustic model. In this context, the English diphtongs /OI/ and /eI/ were added to the dictionary.

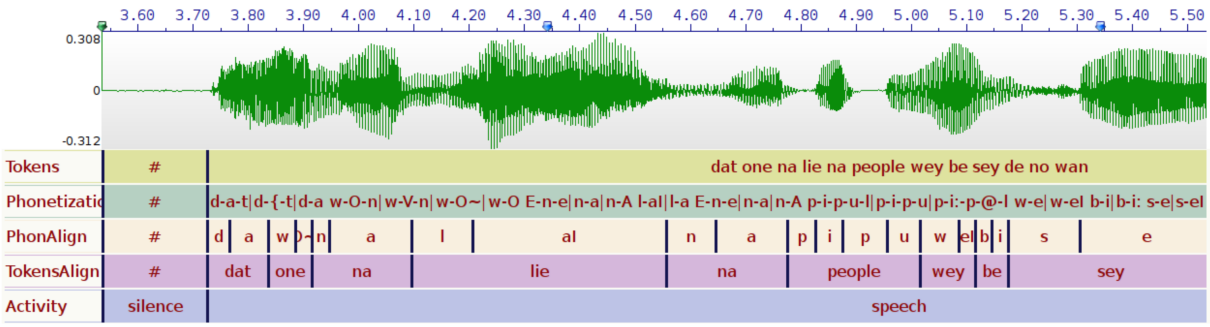

Figure 1: Example of result of the automatic annotations of Naija

4.2. Acoustic model

Acoustic models were created using the HTK Toolkit (Young and Young, 1993), version 3.4. The models are Hidden Markov models (HMMs). Typically, HMM states are modeled by Gaussian mixture densities whose parameters are estimated using an expectation maximization procedure. Acoustic models were trained from 16 bits, 16000 hz wav files. The Mel-frequency cepstrum coefficients (MFCC) along with their first and second derivatives were extracted from the speech in the standard way (MFCC_D_N_Z_0). The training procedure is based on the VoxForge tutorial. The outcome of this training procedure is dependent on both: 1/ the availability of accurately annotated data; and 2/ on good initialization.

Of course, such requirements are difficult to fit in, particularly for under-resourced languages. The initialization of the models creates a prototype for each phoneme using time-aligned data. In the specific context of this study with a lack of training data, this training stage has been switched off. It has been replaced by the use of phoneme prototypes already available in some other languages. The articulatory representations of phonemes are so similar across languages that phonemes can be considered as units which are independent from the underlying language (Schultz and Waibel, 2001). In SPPAS package, 9 acoustic models of the same type - i.e. same HMMs definition and same MFCC parameters, are freely distributed with a public license so that the phoneme prototypes can be extracted and reused: English, French, Italian, Spanish, Catalan, Polish, Mandarin Chinese, Southern Min.

To create an initial model for Naija language, all the phoneme prototypes of English language were used. For the missing ones, the nasals /O∼/, /a∼/, /e∼/, /i∼/ and /u∼/ were picked off Southern Min language, and /E∼/ was extracted from French language by using the /Uˆ/ prototype. The vowels /a/ and /e/ were extracted from the French mode; /O/ and /o/ from the Italian one. The following fillers were also added to the model in order to be automatically time-aligned too: silence, noise, laughter.

This approach enables the acoustic model to be trained by a small amount of target language speech data (Le et al., 2008). The Naija model was created by using the 8 files described in Table 1. Of course, as soon as transcriptions of recordings are available, such new data will be introduced in the training procedure and the model will be updated.

5. HLT tools

5.1. Automatic tokenization and phonetization

In recent years, the SPPAS software tool has been developed to automatically produce annotations, including the alignment of recorded speech sounds with its phonetic annotation. The multi-lingual approaches that are proposed enabled us to adapt some of the automatic annotations of SPPAS to Naija language. An example of Text Normalization, Phonetization and Alignment of a Naija speech segment is proposed in Figure 1.

Tokenization of the Naija language is very similar to the English one. For the purpose of our multimodal studies, we slightly adapted the Text Normalization (Bigi, 2014) and Phonetization systems (Bigi, 2016). For the text normalization, we had only to add the list of words of the Naija language into the ”resources” folder of SPPAS. Such lexicon was created by retrieving the list of words from the corpus published in (Deuber, 2005), and from the first transcribed files of our corpus. From the orthographic transcription, the text normalization system produces tokens (first line of annotations in Figure 1). These can then be used by the automatic phonetization system (second line of Figure 1). For that purpose, we simply had to copy the pronunciation dictionary of Naija into the ”resource” folder of SPPAS.

5.2. Automatic speech segmentation

Automatic forced-alignment is the task of positioning a sequence of phonemes in relation to a corresponding continuous speech signal. Given a speech utterance along with its phonetic representation, the goal is to generate a time-alignment between the speech signal and the phonetic representation. Experiments in this paper were carried out by adding the Naija acoustic model into the ”resources” folder of SPPAS. The automatic alignment can be carried out either using HTK (HVite) or Julius decoder engines (Lee and Kawahara, 2009). Julius is the default aligner used in SPPAS. It produced the alignment of phonemes and tokens as shown in the 3rd and 4th line of Figure 1.

5.3. Experiments

Some experiments were conducted to evaluate the accuracy of the phoneme alignments. It was evaluated using the Unit Boundary Positioning Accuracy - UBPA that consists in the evaluation of the delta-times (in percent-

age) comparing manual phonemes boundaries with the automatically aligned ones. Obviously, the main acoustic model can't be evaluated because all the available data was used to train the model. However, we performed some experiments to have a quick glance at the accuracy of the alignments.

An initialization model was created only from the prototypes already available in the other languages, i.e. without using any Naija data nor training procedure. UBPA of such model is 88.36% in a delta-time of 40ms. This first result confirms the suitability of a cross-lingual approach to create an acoustic model.

The other experiments were performed using the leave-one-out algorithm: 8 models were then created. Each model was trained on 7 of our files, and the model was used to time-align the remaining file. The resulting UBPA is then 91.35%, with a detailed result in Table 4. Of course, introducing Naija manually created data into the training procedure increases significantly the accuracy, even if such data are only 3min29sec long. Accuracy of the model will be enhanced with the data currently under construction.

Delta=T(automatic)-T(manual)	Count	Percentage
-0.030 <= Delta < -0.040	39	1.67%
-0.020 <= Delta < -0.030	69	2.95%
-0.010 <= Delta < -0.020	146	6.25%
0 <= Delta < -0.010	525	22.47%
0 < Delta < +0.010	511	21.88%
+0.010 <= Delta < +0.020	510	21.83%
+0.020 <= Delta < +0.030	245	10.49%
+0.030 <= Delta < +0.040	89	3.81%
-0.04 <= Delta < 0.04	2336	91.353 %

Table 4: UBPA of Naija automatic alignment.

UBPA is a unique measurement suitable to get a quick idea of the accuracy of a model or to compare the quality of several models. However, phoneticians often prefer a more detailed evaluation, as we propose in Figure 2 and 3. We easily observe that the automatic system is reducing the duration of the vowels except for /a/, mainly because the beginning of the vowels occurs later that the expected one.

6. Conclusion

This paper presented the first linguistic resources for the Naija language. It is shown that they are useful for HLT tools: it made Text Normalization (including a tokenizer), Phonetization and Alignment automatic annotations available for Naija. These resources will be gradually improved and updated as the project progresses. The lexicon, the pronunciation dictionary and the first acoustic model are all freely distributed into SPPAS since version 1.9. Future development of automatic tools will focus on an automatic syllabification system: syllables will be the unit of further analysis for all instrumental acoustic measurements.

7. References

Bigi, B., 2014. A multilingual text normalization approach. *Human Language Technology Challenges for Computer Science and Linguistics, LNAI 8387*:515–526.

Bigi, B., 2015. SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, 111–112:54–69.

Bigi, B., 2016. A phonetization approach for the forced-alignment task in SPPAS. *Human Language Technology. Challenges for Computer Science and Linguistics, LNAI 9561*:515–526.

Bigi, B., P. Péri, and R. Bertrand, 2012. Orthographic transcription: which enrichment is required for phonetization? In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uur Doan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation*. Istanbul, Turkey: European Language Resources Association (ELRA).

Deuber, D., 2005. *Nigerian Pidgin in Lagos: Language contact, variation and change in an African urban setting*. Battlebridge Publications.

Esizimetor, D.O. and F.O. Egbokhare, 2012. Naija. hawai university web site. language varieties. `http://www.hawaii.edu/satocenter/langnet/definitions/naija.html` (4 July, 2014).

Faraclas, N.G., 1989. *A Grammar of Nigerian Pidgin*. Ph.D. thesis, Berkeley University of California.

Le, V.B., L. Besacier, S. Seng, B. Bigi, and T.N.D. Do, 2008. Recent advances in automatic speech recognition for vietnamese. In *International Workshop on Spoken Languages Technologies for Under-resourced languages*. Hanoi, Vietnam.

Lee, A. and T. Kawahara, 2009. Recent development of open-source speech recognition engine julius. In *Asia-Pacific Signal and Information Processing Association*. Annual Summit and Conference, International Organizing Committee.

Lewis, M.P., F.S. Gary, and D.F. Charles, 2015. *Ethnologue: Languages of the world*. Dallas, Texas: eighteenth edition.

Onyenwe, I.E., 2017. *Developing Methods and Resources for Automated Processing of the African Language Igbo*. Ph.D. thesis, University of Sheffield.

Schultz, T. and A. Waibel, 2001. Language-independent and language-adaptive acoustic modeling for speech recognition. *Speech Communication*, 35(1):31–51.

Wells, J.C., 1997. Sampa computer readable phonetic alphabet. *Handbook of standards and resources for spoken language systems*, 4.

Young, Steve J and Sj Young, 1993. *The HTK hidden Markov model toolkit: Design and philosophy*. University of Cambridge, Department of Engineering.

8. Acknowledgements

This work was financed by the French "Agence Nationale pour la Recherche" (ANR-16-CE27-0007).

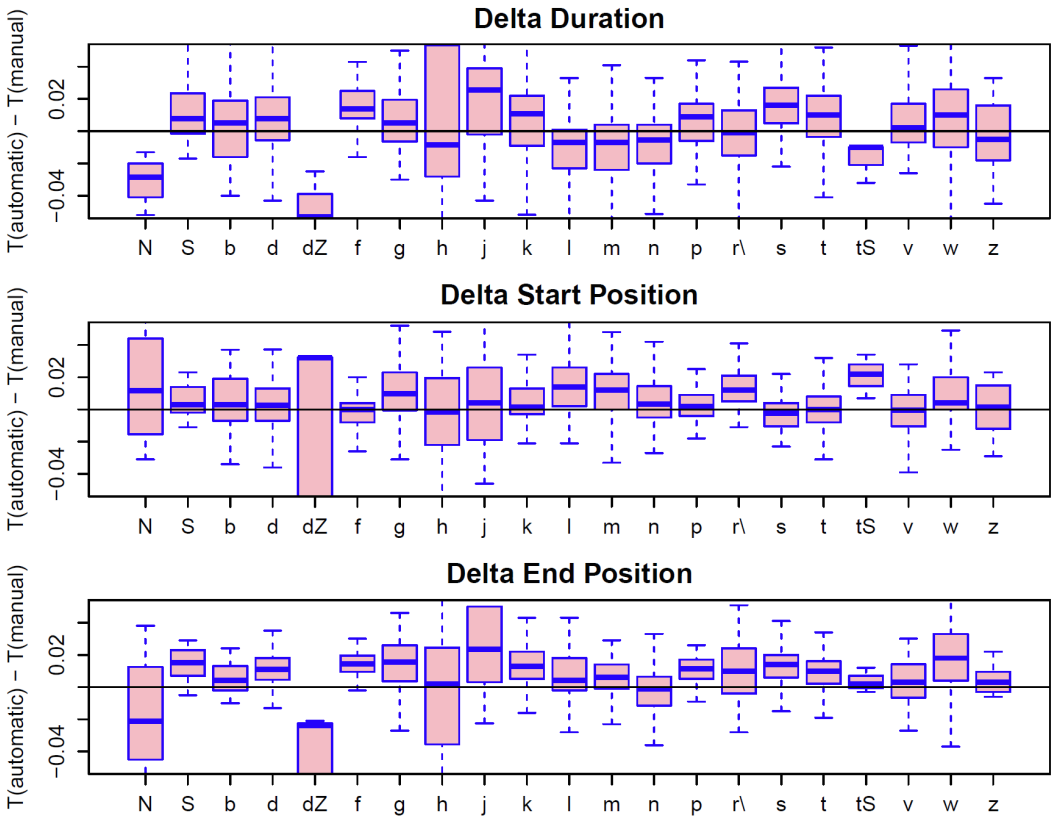

Figure 2: Detailed results of Naija automatic alignment of consonants

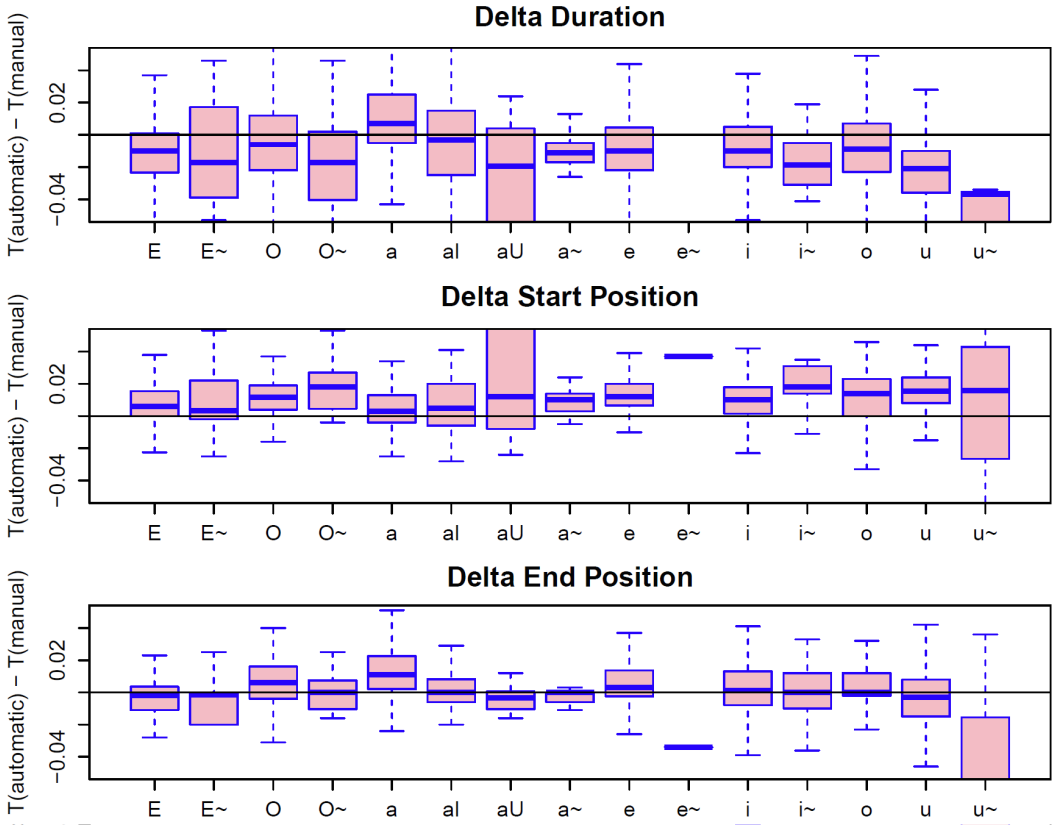

Figure 3: Detailed results of Naija automatic alignment of vowels