Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 7 – No. 36, June 2021 – www.caeaccess.org

Development of a Prosodic Read Speech Syllabic Corpus of the Yoruba Language

Akintoba Emmanuel Akinwonmi
Department of Computer Science
The Federal University of Technology, Akure
Ondo State, Nigeria

ABSTRACT
Literature revealed that the need for annotated database of speech text or audio files is justified primarily by the requirements for corporal entities to conduct basic Natural Language Processing (NLP) studies on a language. Such investigationstraverse thephonetic, aural and etymological representations of the language. Moreover, research of interest can also span grammatic, semantic, pragmatic and syntactic characterizations of the particular language. At a secondary level an annotated speech corpus is desirable for the purpose of speech synthesistypified by Text-to-Speech (TTS) and recognitionas in Speech-to-Text (STT). Yoruba language, a resource scarce language with a wide usage,has sparse andscarce digital resources and its computerization poses unique challenges. Annotated speech corpus is one of such resources.Hence, this research was motivated by the need to contribute to the scanty resources for the language. This research minedtextual inputs from four sources including two Standard Yoruba (SY) fiction, an SY grammar textbook and an SY Online Scripture. A hybrid of Falaschi scheme and the add-on procedure of Radová and Vopálka were applied to extractphonetically balanced text bag of 7376 phrases and sentences with a view to minimizing the extraction cost, while maximizing phonetic coverage of all standard Yoruba syllabic events. The selected text was read by an expert and recorded in a suitable environment and saved as wave files. The wave files were annotated with Praat. A relational database was developed to host the corpus metadata. The corpus performed impressively when tested with a Standard Yoruba TTS. This paper presents the design, implementation, results and other useful information about the research.

General Terms
Speech Processing, Natural Language Processing, Resource Scarce language Processing, Yoruba Language processing

Keywords
Speech, Corpus, Yoruba Language, chunk, Syllabification

1. INTRODUCTION
The Yorùbá language is natively spoken by around 22 million people. It is one of the four major African spoken languages and one of the three major local languages together with English in Nigeria. Standard Yorùbá (SY) is a member of the dialect group used in written Yorùbá in schools and spoken by newsreaders. A speech corpus may primarily support researches at establishing differences between speakers, such as gender, age, environment, speech quality and channels of data. These factors control the speech specificity. Thus, a speech corpus is indispensable in speech processing. It is actually a veritable research material in speech paradigm.

A speech database is essential in the development of systems that support speech processing, synthesis and speech recognition. Thus, fashioning an efficient schema, management protocol, organization, flexibility and monitoring the size of the database is key to have qualitative data-driven speech processing systems. Hence, in designing a speech corpus, the scope of coverage of the language structure (speech units, syntax, word order and grammar) and its phonetic aspect is critical [21].

2. THE YORÙBÁ LANGUAGE
Yorùbá belongs to the Kwa language subgroup of the Niger-Congo languages phylum [1]. It seems that the language has a moderate diaspora. However, its diaspora is wide. It stretches from southwestern Nigeria to Central Togo, East Central part of Republic of Benin, Sierra Leone (Aku variant) to the Caribbean and covering some other islands along the southeastern United States coast. All through the area, the dialects intermix with each other, with Spanish and French forming Gullah variant in the U.S.A and Nagô variant in Brazil. Many slaves brought to North America during the 18th and 19th centuries were of Yorùbá -speaking origin [2].

The language is natively spoken by around 22 million people.It is one of the four major African spoken languages and one of the three major local languages together with English in Nigeria. The highest concentration of its global speakers is in Lagos, Oyo, Ogun, Ondo, Ekiti, Osun, Kwara some part of Edo, and Kogi states in Nigeria. As reported in [3], Yoruba dialect groups include:

a. North-West Yoruba (NWY). Spoken in Abẹokuta, Ibadan, Ọyọ, Ọgun and Lagos (Eko) areas
b. Central Yoruba (CY) Spoken in Igbomina, Yagba, Ifẹ, Ekiti, Iworoko Ekiti, Akurẹ, Ẹfọn, and Ijẹbu areas.
c. South-East Yoruba (SEY) Spoken in Okitipupa, Ondo, Ọwọ, Ikare, Sagamu, and parts of Ijẹbu.

2.1 Yorùbá Phonology
Standard Yorùbá (SY) is a member of the dialect group used in written Yorùbá in schools and spoken by newsreaders. It originated in the 1850s after [4]. Though, it is largely based on the Oyo and Ibadan dialects, it incorporates several features from other dialects [5].

Segmental phonology of Yorùbá alphabets includes seven vowels (a, e, ẹ, i, o, ọ, u), 18 consonants (b, d, f, g, gb, h, j, k, l, m, n, p, r, s, ṣ, t, w, y) and 5 nasalized vowel phonemes (an, ẹn, in, ọn, un). Yorùbá is tonal (accented) with 3 contrastive tones and 2 allotones. The Yoruba letters without diacritics resemble the International Phonetic Alphabet equivalents in pronunciation. It differs on the labial-velarstops [k͡p] (‹p›) and [g͡b] (‹gb›). Both consonants are pronounced concurrently.

Moreover, the glyph underneath vowels in open vowels constricts pronunciation with ‹ẹ› pronounced as [ɛ̯], ‹ọ› as [ɔ̯], ‹ṣ› as [ʃ], ‹y› like the English y, and ‹j› as [dʒ]. Nasalized vowels are specified as oral vowels followed by "n".

2.2 Tone and Intonation in Yorùbá

Tonal configurations ofYoruba alphabets are the following:

A, a, Á, á, À, à, B, b, C, c, Ç, ç, D, d, E, e, É, é, È, è, Ê, ê, Ë, ë, F, f, Gb, gb, G, g, H, h, I, I, Í, í, Ì, ì, J, j, K, k, L, l, M, m, N, n, O, o, Ó, ó, Ò, ò, Ô, ô, Ö, ö, Õ, õ, P, p, R, r, S, s, ß, ÿ, T, t, ª, Þ, þ, Ý, ý, U, Ú, ú, Ù, ù, W, w, Y, y

Three diacritics that are used on vowels and syllabic nasal consonants depict tones. The language allows more than one tone per syllable. The following different syllabic vowel forms are acceptable:

À, Ã/A, Á, È, Å/E, É, Ê, ¦/ẸË, Ì, ¨/I, Í, Ò,/O, Ó, Õ, ±/Ô, Ö, Ù, Ø/U, Ú,

à, ã/a, á, è,å/e,é,ê,µ/e,ë,ì,·/i,í,ò,»/o,ó,õ,¾/ô,ö,ù,ø/u,ú.

In SY, syllabic tones differentiate words of identical segmental phonemes. There are three tonal phonemes H(igh), M(id), and L(ow) [4] Corresponding to an acute accent ‹´›, optional macron ‹¯› and grave accent ‹`›. H tone occurs in word-initial position only in marked consonant-initial words. The following monosyllables can have different meanings as shown:

di: dí (H) to block; dì (L) to pack/tie and dī (M) be deafened

Disyllabic tonal strings in Yoruba give: HH, HM, HL, MH, MM, ML, LH, LM, LL, a total of Tx configurations from T tones and x syllables with less than these number of acceptable (meaningful) configurations in most words [6]. Others are rather nonsensical as this type of parsing only considers syntactic factors; semantics is completely ignored.

Table 1. Tonal Configuration of Disyllabic Yoruba Words

Tones	Word	Meaning	Word	Meaning	Word	Meaning
MH	āyé	World	īyá	a plant	ẹ̄gbá	*
MM	Āyē	A Town	īyā	*	ẹ̄gbā	cudgel
ML	Ayè	A Town	īyà	*	ẹ̄gbà	*
HH	Áyé	*	Íyá	*	ẹ́gbá	*
HM	áyē	*	íyā	*	ẹ́gbā	*
HL	Áyè	*	Íyà	*	ẹ́gbà	*
LL	Àyè	Space	Ìyà	Penance	ẹ̀gbà	Bangle
LH	Àyé	*	Ìyá	Mother	Ẹ̀gbá	A Tribe
LM	àyē	*	ìyā	*	ẹ̀gbā	*

* Unattested

Most words start with a vowel, which is Low or Mid but not High. In SY, tones occur freely in lexical representations, without apparent restrictions on word melodies [1].So, there are 3 possible tonal patterns for monosyllables, 9 patterns for disyllables, and so on. Moreover, in Yorùbá and other syllable-timed languages such as Spanish, Greek, French, Italian, and Telegu, syllables are periodic and often end with a vowel. Whereas in stress-timed languages such as English, Arabic, Russian, and Thai syllable, stress recur approximately periodically and most syllables end with a consonant. Tone in Yorùbá language has been adequately researched and well documented [7][8]. However, the distinctive intonation system due to Yoruba tones continues to be inadequately researched. To this, [9] reasoned that the distraction of linguists' attention from the suprasegmental features of intonation by the impetus on the semantic role of tone is partly responsible for the trend. In spite of advances made in investigating the prosodic features of major world languages, Yoruba has not received adequate research.

2.3 Prosody in Speech

In speech, the acoustic changes in frequency, amplitude, and duration that form grouping, prominence, and intonation are called "prosody"; these features are systematically derived within each language and are thought to be rule-governed and distinct from other structural levels of linguistic analysis [10].

The role of prosody in natural speech is very crucial. The resultant effect of the variations in speech parameters that is specified by changes in duration, pitch and loudness of speech units is perceived as prosody. In synthetic speech, the prosody must be made as similar as possible to that of the natural in order to have a high level of speech intelligibility and naturalness.

Prosody is responsible for the major part of the intelligibility and naturalness of spoken sentences. Syntactic, semantic, as well as emotional information are all conveyed via prosody. Prosodic aspects are often divided into features such as in English stress and intonation. Prosody is in short can be described as the "music" of speech [11].

2.4 Nature of Yoruba Prosody

Different approaches have been used for generate Yoruba prosody ranging from the use of some simple rules to the use of statistical approaches. Such statistical approaches include artificial neural networks, CART, probabilistic approach and hybrid methods. Yoruba is a syllable-based tonal language as such its prosody is characterized by the following properties:

a. Each Yoruba character is pronounced as a syllable.

b. There is a neutral tone called middle tone, a low and a high typical tone. Similar simple syllable types carry different functional meanings if they have dissimilar tones.

c. In terms of prosody, an utterance segments into minor prosodic units including words, minor phrase, and intonation phrase.

d. Each syllable of a Yoruba utterance has a tonal individuality while the utterance exhibits a global intonation.

Considering the foregoing, Yoruba pitch contour has two parts: a restricted component for tone and an overall component for the utterance's general intonation. Thus, the final pitch contour of an utterance is an outcome of the resultant effect of the two components.

2.5 Text Decoding

Text decoding in Text-To-Speech (TTS) is a classical disambiguation problem which entails the determination of the only one correct sequence of words which gave rise to the text sequence of words in a tokenized text. With numbers, dates, named entities and other non-natural language text there may be a few diverse appropriate word chains. Text coding disambiguates such different word chains. With regards to a well performed part-of-speech (POS) tagging, the English words: read, present, record each can have two different pronunciations, and so this is the type of ambiguity needing resolution. Another case is how resolve ambiguity of the language's verbal component, for example, knowing whether 2021 is a number or a year. Resolving this involves: determining the semiotic class for each word or token; resolve any ambiguity for each chain of word/tokens in each semiotic class, then derive their underlying forms and convert each non-natural language classes (dates, numbers and named entities, cardinal numbers, ordinal numbers, telephone numbers, money, percentages and measures) into correct verbalization sequence.

In general, words with alternative spellings, abbreviations, unknown words, homographs (accidental homographs, part-of-speech homographs and abbreviation homographs) and mis-spellings require a sophisticated text decoding models to handle. However, this research will be limited to developing a prosodic TTS for natural language texts. Fair enough, [12] showed that the orthography of SY was shallow. Thus, synthesizing speech from Yorùbá text may not need complicated syntactic analysis. Unambiguous diacritical characterization in SY orthography clearly eradicates the vagueness in pronunciation of otherwise synthetically clumsy words thus a POS algorithm as in [13]is not necessary.

2.6 The Need forSyllabification

[23] categorized Yoruba as a resource-scarce language. A Resource-scarce language is one which has scarce or few digital resources and its computerization poses unique challenges. These challengesinclude unavailability of electronic lexica, standardized electronic corpus and tools such as Part-of-Speech (POS) tagger, phonetiser, syllabificator and optical character Recognition (OCR) tool. It also includes the lack of the political and the financial strength to cause developers and decision-makers in the computer industry to give attention to the needs of the language.

Computerization of Yoruba is more compounded by the absence of basic text processing tool like a well-suited Yoruba keyboard and optical character recognition (OCR) in spite of its large language sphere. Optical character recognition is a technology that enables the conversion of different types of documents such as scanned paper document, PDF files or images captured by a digital camera into editable and searchable data. Availability of which will ease the initial process of text acquisition for computational or research purpose.

A vital challenge encountered when developing an automated syllabificator for TTS system or a speech synthesis system is posed by the need to build a proper speech corpus for the necessary system training and testing. The main hurdles to cross in doing these involve the need to decide the source of materials and the conversion of the text materials into editable form. The last can be solved in two steps by selecting a set of appropriate sentences from a suitable text source. Thereafter, the selected sentences will be broken into syllables and error encountered as a result of text conversion by the scanner, can be manually corrected.

2.6.1 Application of Language Syllabificator

A language syllabificator provides grammarians, lexicographers, Natural language processing (NLP) experts and other interested parties with better descriptions of a language. It plays a crucial role in automatic speech recognition, text-to-speech systems, and corpus statistics.

2.6.2 Advantages of Using A Language Syllabificator

The syllabified texts in the corpus are time and money saving with access to a much larger corpus than a single individual researcher could construct. A further advantage to working with a corpus in the public domain is the possibility of building on the work others have done using the same corpus or using prior results as a benchmark for testing new research methods.

2.7 YorubaOK Font

This font is aimed at promoting the reading, speaking and writing of the Yoruba language properly. It is complete with all the required intonation and diacritic marks of Yoruba language. It was developed by Mr. James Kass and this development has enabled the typing of Yoruba text properly. YorubaOK Font has a File size of 3375kb. It can work on operating systems like windows XP/Vista/7/8. It is a free source and can be downloaded online [13]. The only way to write Yoruba without any limitations and with universal compatibility is for the proprietary fonts used to conform to the UNICODE standard for Yoruba text. It makes it possible for users to acquire and install a capable file that will then work like a standard font e.g. Arial, Courier or Times New [13].

3. REVIEW OF RELATED RESEARCHES ON STANDARD YORUBA NLP RESOURCES

[14] developed Text Statistics Tool Box that attempted among other tasks to count the number of syllables in a characteristic list (that is a sentence) using English syllable rules. The rules utilized in constructing this toolbox made accuracy difficult as the system can easily misclassified syllables. Another set of efforts at counting the syllables in an English sentence were also made. None of these is guaranteed to give very good results and none was designed with the intent of producing tokens.

The OAK System was developed by the Proteus Project of the Department of Computer Science of the New York University as a "total analyzer for English". The tokenizer tokenizes into words or word forms but not into syllables. The tokenizer in the Stanford Natural Language Processing (NLP) Toolbox of the Stanford NLP Group also tokenize into words and word forms. The English language tokenizer is a Finite State Machine tokenizer based on hand-written rules. The Open NLP tools are a set of Java-based NLP tools that perform sentence detection and tokenization among other tasks for two European languages and one Asian language. The tokenizer which is for tokenizing sentences into words followed the token style for the English Penn-Treebank. NL Tokenizer for English Language tokenizes a text into paragraph, sentences, words (including simple word-form, abbreviations and numbers) that can work in standalone mode or as part of another module.

[15] is closely related to this work. The authors developed a Swahili Text-To-Speech application that has a syllabification module (SM) embedded. The SM takes in a text consisting of strings of Swahili words and output string of syllables to be

fed as input into a Digital Signal Processor that synthesizes the speech also used a rule-set however simpler due to the simpler syllable structure of Swahili.

[16] reported on the development of a syllabicator for Yoruba language. "Syllabicator" is a synonym for "Syllabificator". Their work was motivated by the need to contribute to NLP resources for Yoruba language. They identified a syllabicator as one of the needed research resources for Yoruba as a tonal language. The authors used a declarative rule-based approach for the syllabification of Yorùbá words. The declarative rule-based approach is one in which just one rule that is first deterministically applicable is selected while others are ignored as they do not apply to the given situation. Their method initially determined the length of the selected word from a sentence and thereafter applied a set of rules that takes into consideration the characters in the word and the order or arrangement of the characters within the word as defined by Yoruba orthography. Thereafter the researchers applied 23 rules that were categorized into 4 cases to determine syllable boundaries. Though an evaluated performance of 99.993% accuracy of returning correct syllables from a given word when evaluated on a text size of 90622 words was reported, their system lacks the ability to discriminate against non-Yoruba words. Moreover, some words containing DVn and CVn, especially name identities must be treated specifically. For instance, while their system would syllabify Igbin correctly as I. gbin (2 syllables), Bankole should be syllabified as Ba.n.ko. le (4 syllables) not Ban.ko. le (3 syllables). Ditto for Banke, Banro, kengbe, and other similar SY words as would be the case. Several examples of such Yoruba words exist.

[17] proposed a prosody realization in Text-to-Speech (TTS) applications for Yoruba language. Their work was motivated by the need contribute NLP research tools to Yoruba Language. The objectives of their research were to model prosody from the speech of a Yoruba language speaker and also to develop a Hidden Markov Model (HMM) based TTS system for Yoruba text. Although their proposal underscored the prominence and indispensable of language corpora in Natural Language processing, one major challenge emphasized in their work was in the development of a speech corpus of Yoruba language. In overcoming this challenge an accurate, language specific OCR and Syllabificator is inevitable.

[18] presented Part-of-Speech tagging of Yoruba Standard; Language of Niger-Congo family. Their work was motivated by the need to obtain training data for part of speech taggers and to contribute basic application resources towards Yoruba Language Processing. Their paper aimed to design a Yoruba corpus. The tagging was performed with Support Vector Machine Tool one of the Part-of-Speech taggers widely used. The corpus with 312,562 words, formed from the Web, was annotated with an accuracy of 98.04%. This annotated corpus might be used in translation system. First, the researchers made the choice of tools and then moved on to the Part of Speech Tagging. The first tool of their work is Yoruba language. The authors classified Yoruba as a tonal language which has three tones: high tone (represented by the acute accent), the low tone (represented by the grave accent) and the average tone (represented by the absence of accent).They first performed the text pre-processing (segmentation, removing duplicates) using scripts written in Perl (Perl 5) before performing part of Speech tagging. The researchers considered Part of Speech Taggers as Natural Language Processing tools that are neutral, because they are able to adapt to any language, as long as the corpus and models are available. They used machine learning methods. These methods allow in the first time, to learn the language based on some relevant examples of sentences which are provided. The main result of their research is the creation of the annotated corpus with Part of Speech Tagging for the Yoruba. Moreover, they based their research on a somewhat contestable premise that, Yoruba had a unique alphabet containing thirty (30) letters composed of 12 vowels: i, e, ẹ, a, o, ọ, u, in, ẹn, ọn, un, an and 18 consonants: b, d, f, g, gb, h, j, k, l, m, n, p, r, s, ṣ, t, w, y. The result of their research cannot be safely applicable to SY because truly'in', 'ẹn', 'ọn', 'un', and 'an' are not considered as alphabets in Standard Yoruba.

[19] improved on the model of [16]by introducing a language filter made up of a regex (wild character structure) filter pass and/or word-list pass which allowed the system to discriminate against non-Yoruba words. However, the underlying rules and design became more complex even as the system would still not be able to handle some words containing DVn and CVn, especially name identities that must be treated specifically (Bankole would still be wrongly returned as Ban.ko.le - 3 syllables).

4. SYSTEM DESIGN

Concatenate synthesis and formant synthesis are the predominantly used approaches for producing speech waveforms. The choice of which of the technologies to use depends on the intended uses of a synthesis. They both have varying strengths and weaknesses. This research applies the concatenate-synthesis-based unit selection using syllables as the base units.

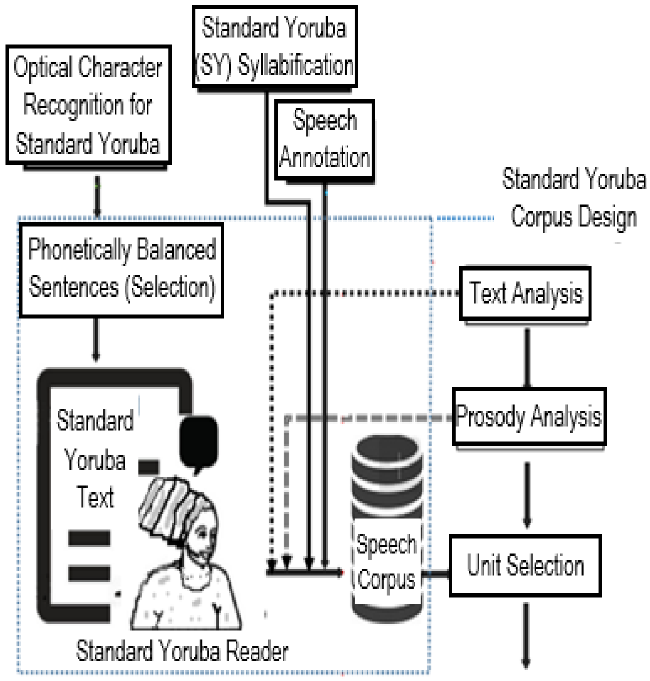

Figure 1: Standard Yoruba Corpus Design

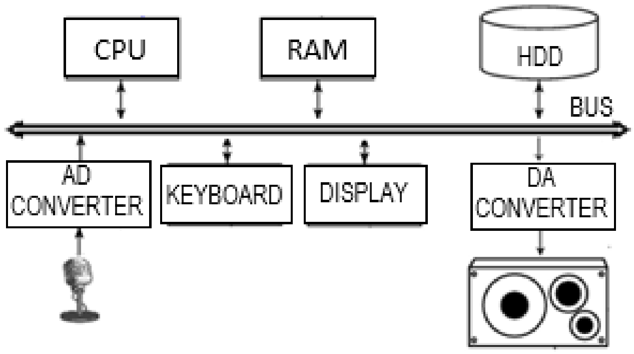

Figure 2: A block diagram of hardware support for Standard Yoruba Corpus

4.1 Design of the SY Read Speech Corpus

The need for a read speech database is justified firstly by the requirements for corporal entities to conduct basic research on a language as posited by [20]. These can span the auditory, phonetic, etymological, semantic, syntactic representations of the particular language. This factor concerns the design of the database: it also affects the response time in speech systems. Moreover, a speech corpus may primarily support researches at establishing differences between speakers, such as gender, age, environment, speech quality and channels of data. These factors control the speech specificity. Thus, a speech corpus is indispensable in speech processing. It is actually a veritable research material in speech paradigm. A speech database is essential in the development of systems that support speech processing, synthesis and speech recognition. Thus, fashioning an efficient schema, management protocol, organization, flexibility and monitoring the size of the database is key to have qualitative data-driven speech processing systems. Hence, in designing a speech corpus, the scope of coverage of the language structure (speech units, syntax, word order and grammar) and its phonetic aspect is critical [21].

4.2 Read Text Corpus Database

This subsection presents the design of the relational schema of the corpus database, built on a conceptual schema design. The Extended Entity-Relationship modeling tool is used to derive the design as shown in Figure 3. The resulting relational database structure is shown in Figure 4.

The relational schema is defined as follows:

a. CORPUSTEXT (TextID, idType, Source, TextName, DateAquired, Language, LangVariant, TextPath)

b. ENVIRONMENT (EnvironID, Description, Remark)

c. READER (ReaderID, SurName, FirstName, OtherNames, Sex, DOB, Address, email, Speciality, OtherTraining)

d. RECORD (RecordID, ReaderID, TextID, TextData, Time, Duration, EnvironmentID, Format, NumChannels, SampleRate, BitRate, BitsPerSample)

e. TEXTTYPE (TextTypeID, TextType, Description)

f. UNIT (UnitID, RecID, UnitTypeID, F0, F1, F2, Duration, Intensity, Gaussian, MFCC)

g. UNITTYPE (UnitTypeID, Description, Category, Remark)

h. CODEBOOK (CodeID, Code, UnitID)

i. WORD (WordID, Word, POS, Meaning)

These are represented as mapped, in the conceptual schema design in Figure 4.The entity types in the read text corpus database include CorpusText, Word, TextType, Unit, UnitType, Reader, Record, CodeBook and Environment. They are all regular entities. The mapping of EER schema to Relational schema firstly involved using the regular (strong) entities type E in the EER schema to derive a relation R that comprises all the simple attributes of E; and then choosing suitable key attributes of E as primary key for relation R. Where multiple keys in E were identified in the ER design, the keys are retained so to serve as secondary (unique) keys of relation R. The primary keys are mainly necessary for maintaining entity uniqueness and indexing functions.

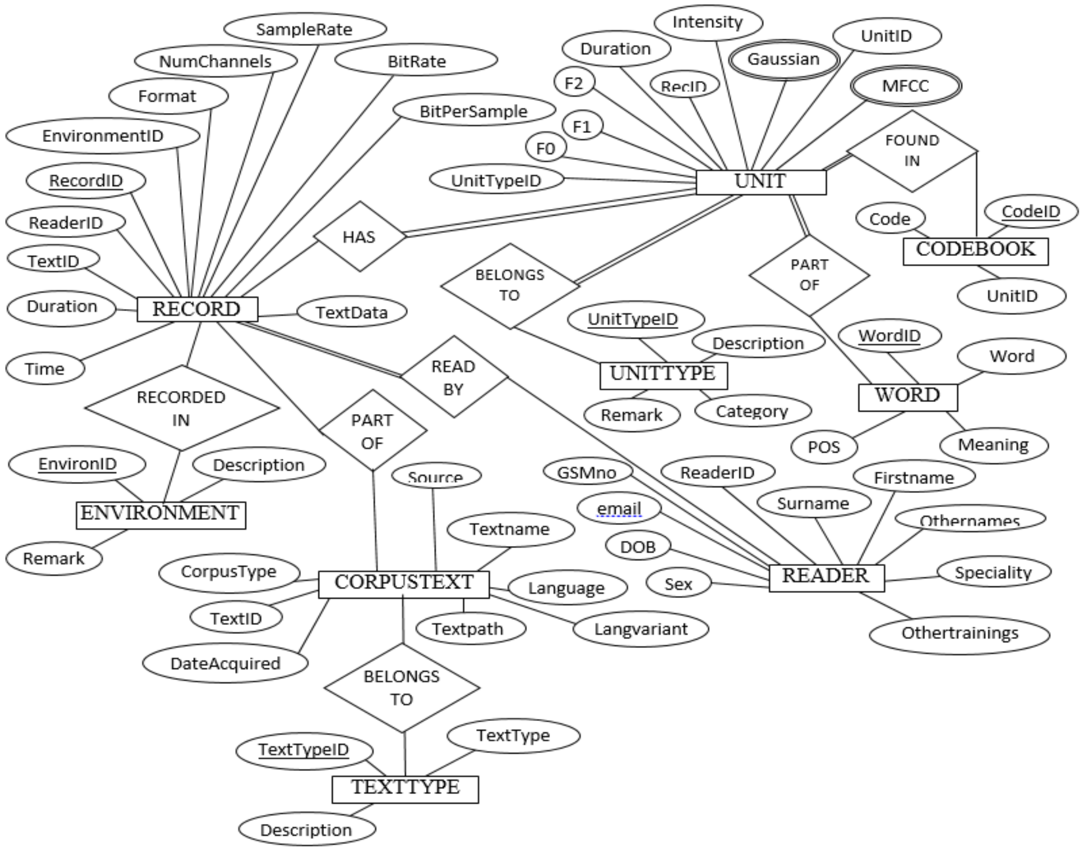

Figure 3: Extended Entity-Relationship model of the SY speech corpus database and its corresponding Relational Database Structure respectively

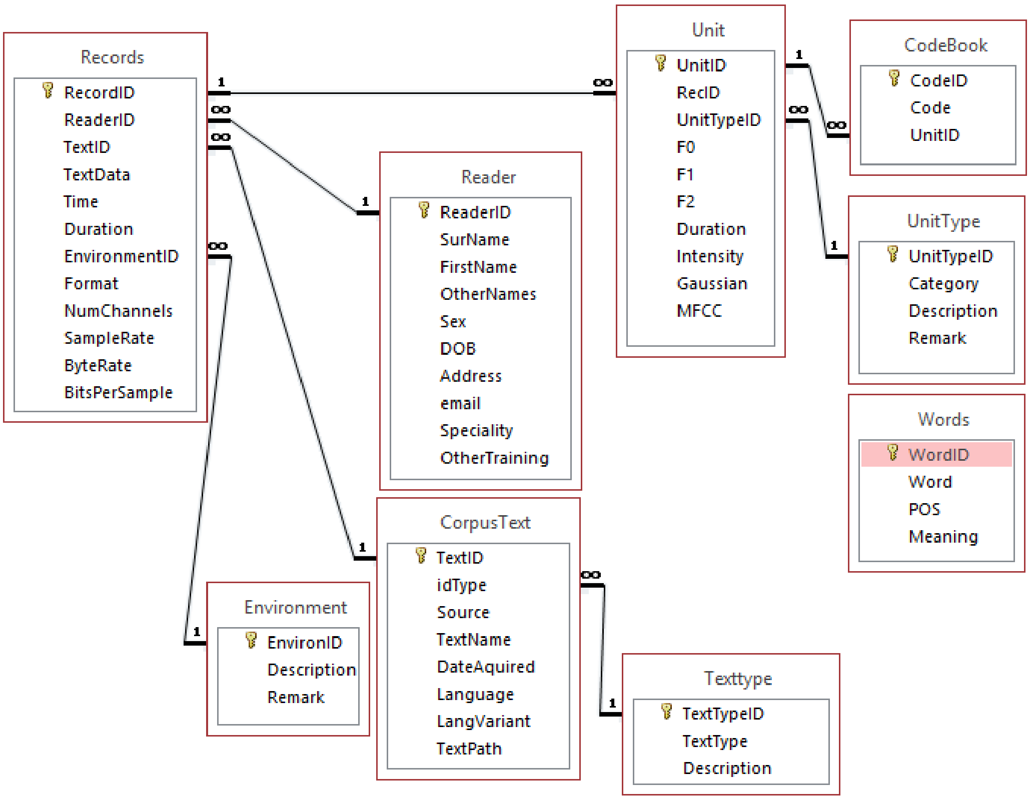

Figure 4: Relational Database Structure of the SY speech corpus database

4.3 Selection of Sentences for the Corpus

A careful selection of sentences for the corpus is necessary. Steps were taken in the method, to make sure that there is a balance in the length and spread of syllabic events in the selected sentences for the corpus. The complexity of syllabic events in the sentences is restricted by relational algebraic expression in (1)

$$\sigma(([\text{NoOfSyllable}]<=15 \text{ AND } [\text{NoOfSyllable}]>=3) \text{ (SENTENCES)} \qquad 1$$

Thus, in the output

$$\exists \{ş_0, ş_1, ş_2 \ldots ş(_{n}-_1)\} \in Ş \text{ s.t. } \{15<= \tilde{n}_i>=3\} \wedge \{n\} \sim \{i\} \qquad 2$$

where n = Number of required sentences, Ş = Set of all sentences, $ş_x$ = sentence of index x, and $\tilde{n}_i$ = Number of events in sentence *i*. S is used to get phonetically balanced sentences from the corpus as follows:

$$S = -\sum_{i=1}^{I}\left|\frac{m_i}{m} - \frac{n_i + n'_i}{n}\right| \qquad 3$$

$$m = \sum_{i=1}^{I} m_i \qquad 4$$

$$n = \sum_{i=1}^{I}(n_i + n'_i) \qquad 5$$

when I = Number of desired phonetic events,
m_i = Number of i^{th} corpus events,
n_i = Number of i^{th} events in the cumulative selection
n'_i = Number of the i^{th} phonetic events in the scanned sentence. Scanning of sentences is repeated.

At each epoch, the sentence with the highest S is chosen for*n* epochs. Thereafter, the selected sentences are shingled and sorted:

For *Sentence* = abcab and *ShingleWidth* =2
Where a,b,c are syllables
Shingles(abcab,2)->{ConCat(a,b),ConCat(b,c), ConCat(c,a)} 6

At each epoch, the sentence with the highest S is chosen for n epochs. Thereafter, the selected sentences are shingled and sorted:

4.4 SY – Syllabification and Shingles () Function

Syllabification (syllabication) is the separation of a word into syllables, whether spoken or written. Thus, in the context of this research, a syllabificator (syllabicator) is the software that can be used to perform syllabification (syllabication). It is a language-based operation.

The syllabificator model proposed for this research seeks to improve over the performance of the syllabicator in [16] and syllabificator in [13]. This model adopts the NLP concepts of alphabetic shingle and valid syllable mappings or structures to derive the valid syllabic tokens from the set of shingles generated from a given word.

NLP models that allot probabilities to strings of words are known as language models (LMs).The N-gram is the simplest model that allots probabilities (likelihoods) to sentences, phrases and series of words. In NLP, the elements of an n-gram can be phonemes, syllables, letters, words or pairs of these base kinds. The n-grams typically are collected from a text or speech corpus. N-gram model is a probabilistic model used for forecasting the next element in such a sequence in the form of a (n − 1)–order Markov model. For simplicity and scalability, n-gram models are currently broadly used in communication theory, probability, biological sequence analysis, data compression and statistical NLP. N-grams are also called Shingles [22]. Shingling is used to convert a document or textual material to unique sets. At the word level it takes consecutive words and group them as a set. A k-gram is a consecutive set of k characters. By so doing, the set of all 1-shingles at the word level is the bag of words model.

Now, let *NGramAl* represent the alphabetic shingling or character shingling function, that is, an n-gram in which the elements are letters within a word. The process is described as follows:

String Array[n]*NGramAl*;

String *NGramAl ;*

NGramAl (theWord, l) ≡ {S0, S1,…, Sn-1}; 7

theNGramAl = [S0, S1,…, Sn-1]; 8

Where *theWord* is the target word;

l is the *NGramAlWidth* (the number of character tokens in each n-gram in the set of all continuous *n* sequences of character-tokens).

Where:$LEN(Si) \leq l$ and $0\leq i\leq n-1$

An n-gram for a document is a sequence of n characters that appears in the document.

Thus, NGram(“abcab”,2) returns {“ab”, “bc” , “ca”}

Likewise, NGramAl (”abcabcd”, 3) returns {“abc”, “bca”, “cab”, “bcd”}.

Optionally, shingle is a bag, hence will return {“abc”, “bca”, “cab”, “abc”, “bcd”} containing two strings of “abc”.

4.4.1 The Sy-Syllabificator Design

The design of the SY-Syllabificator consists of three logical modules or subsystem. These are:

a. SY-Syllabificator File/Sentence Reader module;
b. SY-Syllabificator main module; and
c. SY Syllable Structure Generation module.

The contiguous architecture and functional design of the system is presented in figure 5.

The architecture and functional description of the three parts are presented in the next subsections.

4.4.2 SY-Syllabificator File/Sentence Reader Module

Figure 5 gives a graphical illustration of the process flow of the SY-syllabificator file/sentence reader module. This subsystem responsible for initializing the syllabification process at the beginning and also responsible for terminating the process at when there are no more sentences to read. It starts by reading in the File1 (“*.yor”) Unicode file that contains the sentences been syllabified. Here, the file is delineated into sentences as necessary.

Furthermore, each sentence is read and then processed in a loop of the form:

```
While (File1.CurrentSentence.Trim() > "" )
{
System.File1.CurrentSentence.Syllabify();
```

```
} else {
System.File2.Write (Syllables & Structure);
Stop;
}
```

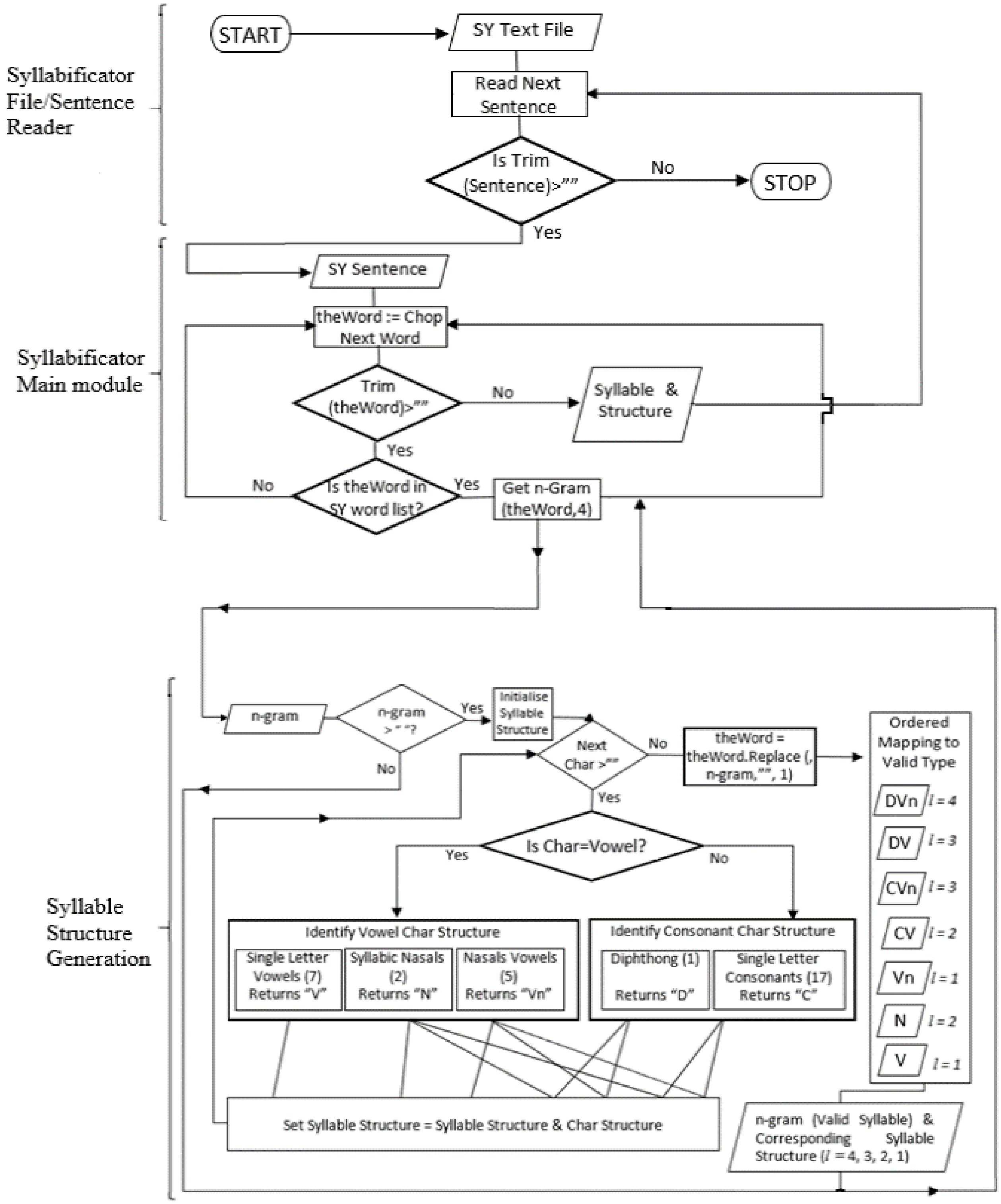

Figure 5: The architecture and functional design of the SY-Syllabificator

The loop exits when the function *File1.CurrentSentence.Trim()* returns an empty string signaling no more sentences.

While the epochs continue, each read sentence is sent to the SY-syllabificator main module for necessary parsing into words and then to the SY syllable structure generation module where further processing over each epoch ultimately returns a duo of SY syllables and their corresponding syllable structures for each given sentence. The pair is written into another file File2 ("*.syl") in Unicode format.

The diacritic transformation is performed and tokenized into words. The word is then parsed to determine whether the structure is Yoruba before finally passed to the syllabication module. For each of the given word-lengths, a set of rules were designed to cover the possible combinations of characters and the syllables that could be formed from them.

The diacritic transformation is performed and tokenized into words. The word is then parsed to determine whether the structure is Yoruba before finally passed to the syllabication module. For each of the given word-lengths, a set of rules were designed to cover the possible combinations of characters and the syllables that could be formed from them.

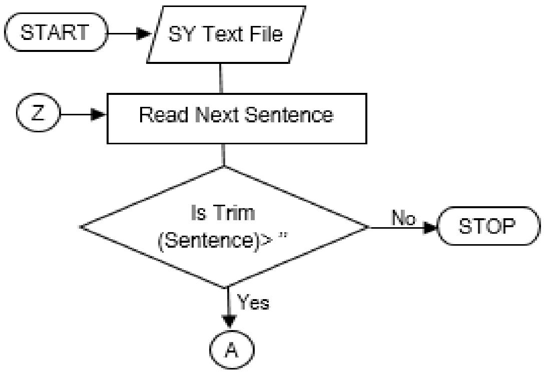

Figure 6: SY-Syllabificator File/Sentence Reader

4.4.3 SY-Syllabificator Main Module

Figure 7 gives a functional graphical representation of the SY-syllabificator main module. This subsystem, on each epoch created in by the SY-syllabificator file/sentence reader module, receives a SY sentence as its input.

Furthermore, it parses the sentence into word tokens. It then creates a loop for each word in which it generates the NGramAl(theWord,4) per word, thereby generating quasi-syllable tokens of widths 1 to 4. Moreover, the subsystem passes each set of NGramAl tokens generated to SY syllabificator syllable structure generation module where syllable structures are generated.

4.4.4 SY-Syllabificator Syllable Structure Generation Module

This is the module that is actually responsible for generating syllable structures and returning valid syllables for a given word and for a sentence over the various loops. The process is illustrated by Figure 8. The corresponding NGramAl tokens for each word generated from the SY-syllabificator main module are processed in this module. The process involves the generation of syllable structures for each quasi-syllable from the previous. Then the generated structures are mapped to a list of seven ordered structure prototypes. These seven represent the valid prototypes allow in SY writing.

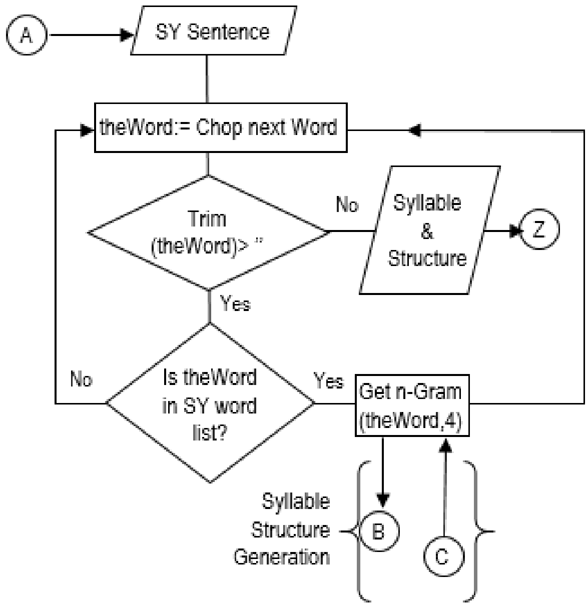

Figure 7: SY-Syllabificator Main Module.

Essentially, at this juncture, the mapping operation returns only valid syllables and their corresponding valid syllable structures while invalid ones are dropped. These are later returned as a duo of SY syllables (σ) and their corresponding syllable structures (σ′) for each given sentence. These processeswere repeated for all words of all sentences in File1 ("*.yor") file. One obvious advantage of this design is that in deciding the validity of each syllable, iteration is done only by mapping to seven syllable structures. These are of V, N, CV, CVn, Vn, DV and CVn patterns.

$$\sigma' \epsilon \{V, N, CV, CVnVn, DV \text{ and } CVn\} \qquad 9$$

This is bound to be faster than the method of [16]and [19] which runs on each given word-length, a set of rules that were designed to cover the possible combinations of characters and the syllables that could be formed from them. Moreover, adaptation of this design to other languages will involve just one step of defining valid syllable structures for the target language. There will be no need to redefine syllabification rules for the new language.

With the model in Figure 8, it is possible to generate equivalent syllabic maps for the syllabic NGrams for each word provided the input is replaced with each Syllabic NGram. Thus, the Yoruba Syllable Mapping Generator modifies to the structure in Figure 8

Where σ = Set ϶ σ|→σ′ ϵσ′

σ =the Set of SY Syllables

σ = a given syllable, an element of σ

σ′ = a given syllable structure, an element of σ′

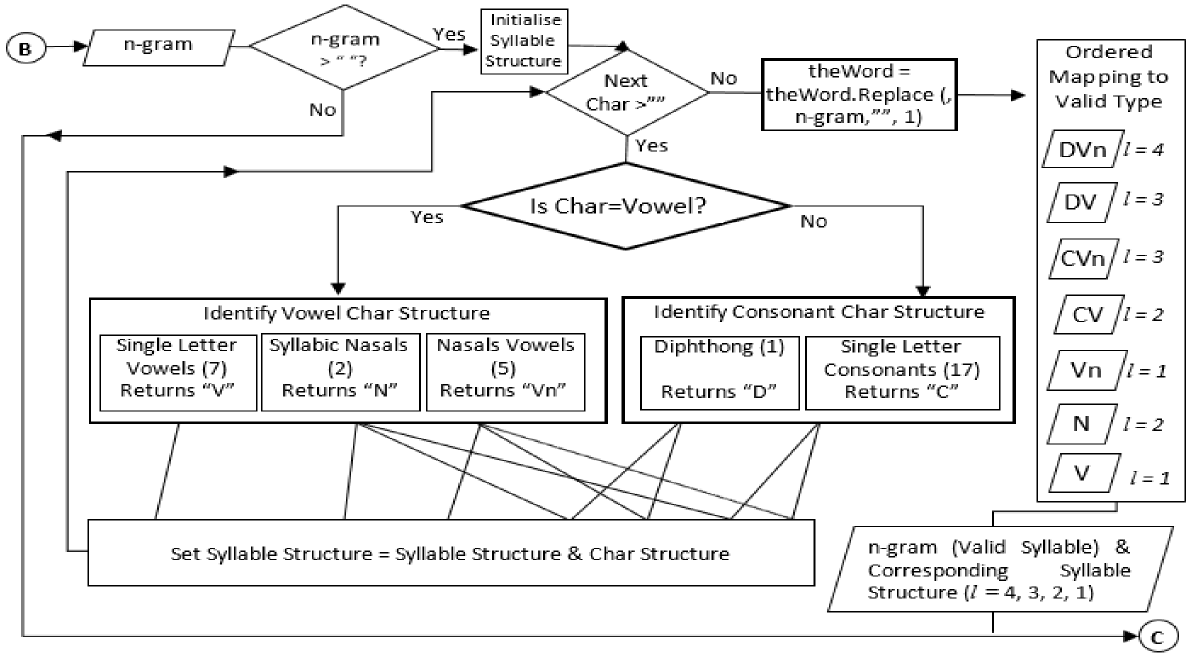

Figure 8: SY-Syllabificator Syllable Structure Generation Module

5. RESULT

A combination of both manual and scripting methods was used in annotating the recorded speech files for the SY TTS.

Segmentation of recorded utterances into words and then syllables was accomplished using Praat toiterating over the recorded speech files and the corresponding texts. Thereafter Praat text grid editor was used to manually correct corporal segmentation errors by listening to each units and visually tweaking unit boundaries on the respective tiers of the text grid. This is a strenuous exercise but necessary to improve the accuracy of the corpus.

5.1 Feature Extraction of the Corpus

Extraction of pitch features for the SY syllables was performed using Praat scripts.

Table 2 depicts the extracted pitch components up to the P10 for certain syllables. Only P0, P1 and P2 were stored on the database in view of the design in Figure 3.

Table 3. Extracted Intensity Features of Certain Syllables in Decibels

Table 4. Extracted Duration Features of Certain Syllables in Seconds

Table 5. Mean Values for Extracted Features of Syllables

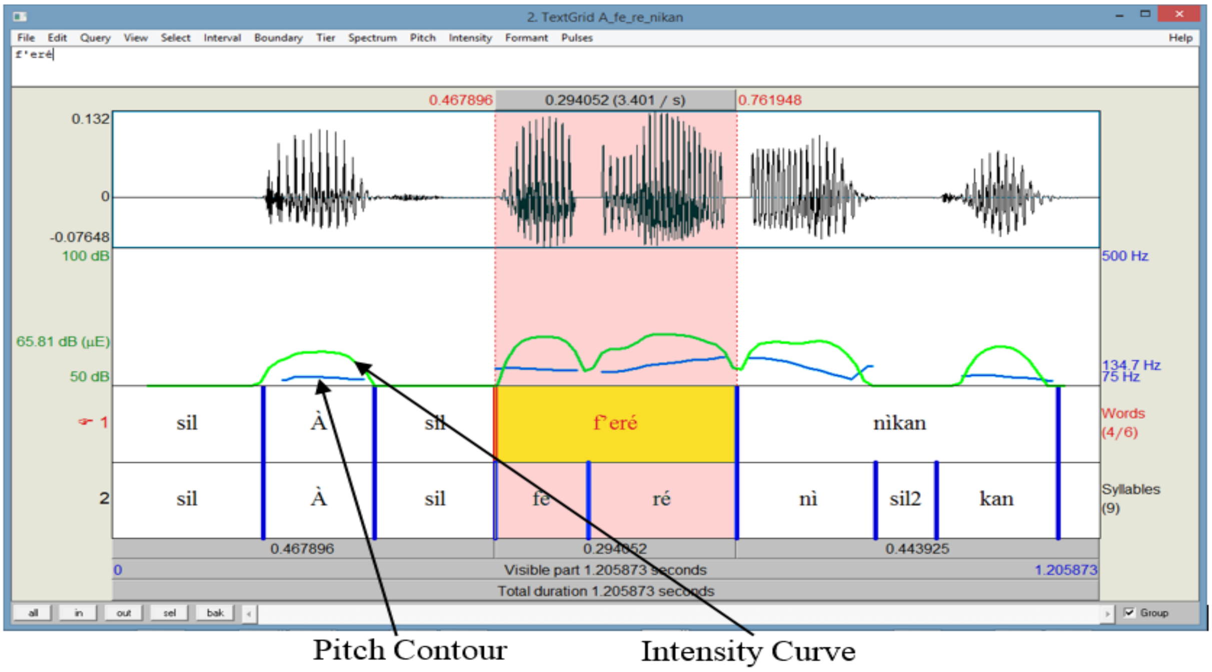

Figure 9: Praat Textgrid Utterance Annotation fora 2-Tier (Words Tier and Syllables Tier)Annotation of *"à f' eré nìkan"*

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 1.2058730158730158
tiers? <exists>
size = 2
item []:
  item [3]:
    class = "IntervalTier"
    name = "Words"
    xmin = 0
    xmax = 1.2058730158730158
    intervals: size = 6
    intervals [3]:
      xmin = 0
      xmax = 0.18503100618199428
      text = "sil"
    intervals [2]:
      xmin = 0.18503100618199428
      xmax = 0.32087027061462775
      text = "À"
    intervals [3]:
      xmin = 0.32087027061462775
      xmax = 0.46789629800053695
      text = "sil"
    intervals [2]:
      xmin = 0.46789629800053695
      xmax = 0.7619483527723553
      text = "f'eré"
    intervals [6]:
      xmin = 0.7619483527723553
      xmax = 1.15508316513033
      text = " nìkan"
    intervals [6]:
      xmin = 1.15508316513033
      xmax = 1.2058730158730158
      text = "sil3"
```

Figure 10: Textgrid Data for *“à f' eré nìkan“* Utterance Showing Words Tier Annotation (Higher Level).

```
File type = "ooTextFile"
Object class = "TextGrid"

xmin = 0
xmax = 1.2058730158730158
tiers? <exists>
size = 2
item []:
item [2]:
    class = "IntervalTier"
    name = "Syllables"
    xmin = 0
    xmax = 1.2058730158730158
    intervals: size = 9
    intervals [3]:
      xmin = 0
      xmax = 0.18503100618199428
      text = "sil"
    intervals [2]:
      xmin = 0.18503100618199428
      xmax = 0.32087027061462775
      text = "À"
    intervals [3]:
      xmin = 0.32087027061462775
      xmax = 0.46789629800053695
      text = "sil"
    intervals [2]:
      xmin = 0.46789629800053695
      xmax = 0.581362036526619
      text = "fe"
    intervals [6]:
      xmin = 0.581362036526619
      xmax = 0.7619483527723553
      text = "ré"
    intervals [6]:
      xmin = 0.7619483527723553
      xmax = 0.9313479060648159
      text = " nì"
    intervals [4]:
      xmin = 0.9313479060648159
      xmax = 1.0064590287510955
      text = "sil2"
    intervals [23]:
      xmin = 1.0064590287510955
      xmax = 1.15508316513033
      text = "kan"
    intervals [5]:
      xmin = 1.15508316513033
      xmax = 1.2058730158730158
      text = "sil3"
```

Figure 11: Textgrid Data for “*À f'eré nìkan*“ Utterance Showing Syllables Tier Annotation (Lower Level)

Table 2. WordsExtracted Pitch Features of Certain Syllables in Hertz

Id	S	P0(Hz)	P1(Hz)	P2(Hz)	P3(Hz)	P4(Hz)	P5(Hz)	P6(Hz)	P7(Hz)	P8(Hz)	P9(Hz)	P10(Hz)
36	Yún	110.2868	114.1006	116.7839	119.5503	122.3146	125.3720	128.8481	132.1298	134.8732	137.7730	140.5459
47	Yún	232.4902	226.5086	218.1436	209.0240	200.0097	191.3129	183.1027	174.5445	165.5585	157.6415	150.4485
72	yun	102.5374	103.8606	105.6650	107.5538	108.9202	110.5460	112.2565	113.9411	115.6384	117.3619	118.3995
108	Yù	96.6793	96.6440	96.5269	96.9954	98.0269	98.3835	98.8485	100.5472	100.3138	101.8138	101.8186

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 7 – No. 36, June 2021 – www.caeaccess.org

119	Yù	143.2035	134.2543	126.4178	119.2112	113.1261	106.4613	98.8049	92.2552	86.7868	81.6931	76.1109
155	Yú	219.3048	212.0127	202.4239	193.3997	185.7468	178.0436	171.7262	165.9273	161.4758	157.8539	154.3605
828	yë	95.8243	96.7083	97.9845	99.5912	101.5147	104.4873	108.2068	112.4411	117.3821	122.9313	128.9399
972	ye	92.6868	92.5907	92.8153	93.7419	94.5734	93.8305	93.4863	94.0978	95.5076	96.7879	98.2397
1008	yçn	93.0858	94.0544	95.5027	96.7725	97.5742	98.5053	100.0424	101.4363	103.3852	105.7821	107.8814
1019	yçn	215.2838	210.6033	202.8377	192.1555	181.8663	172.8903	164.0620	155.5401	147.4894	140.7716	133.9680
2016	wí	118.6305	123.8112	130.9109	137.1851	142.8654	148.6868	156.3495	160.0578	163.6268	166.1155	168.4327
2027	wí	188.5958	186.9798	184.4597	180.2741	175.8855	171.1899	166.7354	160.1373	154.0799	145.0943	138.2646
2171	wë	181.5146	176.2070	170.1165	165.0042	161.0991	158.5885	156.2436	152.4230	147.8677	142.7734	136.8484
2232	wè	87.5485	89.8510	92.3013	95.0728	98.4751	100.7805	102.7393	104.7054	106.3491	106.3615	105.7807
2279	wé	177.5540	172.4199	166.3148	161.8967	157.1944	153.2258	150.4489	147.1709	143.2561	139.2135	134.4185
2315	we	172.5931	166.4979	160.7098	155.1876	149.6868	143.1152	135.3168	127.4450	119.9253	112.5183	105.5440
2376	wç	101.8549	104.1377	106.6572	108.4570	110.1324	110.9517	111.4319	112.5763	113.7578	114.7338	115.2594
2772	ù	95.2252	95.9270	96.4599	97.4349	98.1635	98.3023	98.5159	98.2636	97.5455	96.7111	95.9025
2808	ú	140.6187	145.8414	151.8351	156.7228	160.5586	164.3926	168.5562	172.5605	175.8520	178.4230	180.7496
2819	ú	234.4417	231.1811	224.2543	214.7099	205.3312	196.3984	187.6061	179.3871	171.0912	162.5179	153.1596
3024	tu	114.3018	117.8508	118.6078	118.5431	118.9915	119.1673	119.4314	119.9830	120.9926	121.3890	121.1462
3215	tö	199.1879	189.8251	177.0654	166.9405	158.7295	151.2303	144.0488	137.0716	130.3283	124.4378	118.1805
3240	tô	109.7954	108.9180	108.7385	109.4576	110.4345	111.2468	111.9139	112.6645	112.5695	111.5007	110.0918
3312	tó	146.2979	147.2157	148.9607	151.3603	154.4362	158.2915	162.1161	165.0174	167.7229	171.7020	182.7588
3575	ti	185.0830	178.9860	168.4533	160.4771	152.4052	143.7944	134.9659	126.4339	118.7894	111.5541	104.3106
3827	te	187.1610	176.7660	162.6942	152.4320	140.7850	133.6531	125.6845	118.1189	111.5541	105.2271	99.7080
4043	tà	142.6785	129.1879	119.7930	109.3154	100.4436	89.4442	80.0444	78.0786			
4068	tá	136.5525	136.9234	137.4230	138.5579	140.8469	143.8584	147.1359	151.2560	154.9611	157.5654	161.1818
4079	tá	204.5551	194.8978	182.5991	171.0992	161.6676	154.2606	147.7255	141.4879	135.8092	129.3375	123.0552
4187	sún	231.5180	223.1693	211.7717	201.0198	191.7057	183.4371	175.6478	168.4265	161.7610	154.9418	149.1385
4212	sun	129.8070	128.0456	126.9342	126.7099	126.8509	127.5854	128.7999	130.8564	133.5138	136.5052	139.4013
4223	sun	209.8370	196.2031	180.9702	167.8933	156.2549	146.6017	138.2234	131.7938	125.7905	119.7435	113.9173
4835	sí	214.7763	212.3132	206.3359	198.3963	190.2926	182.2608	173.9031	165.1563	156.4373	150.5354	147.9966
5256	sán	132.8687	134.6006	136.9148	139.8811	143.1361	146.5432	149.4802	152.5981	155.8062	158.7495	163.3063
5267	sán	185.3142	177.5012	168.1271	160.3069	153.2171	147.0936	140.5348	134.1807	127.7895	121.3457	116.5960
5292	san	104.5808	103.7416	103.2944	103.1315	103.2711	103.9493	105.1077	106.1620	106.6439	105.9845	105.4681
5483	rún	193.4823	187.8375	179.8767	173.9368	167.6460	161.2152	154.5212	148.7684	143.2890	137.7543	132.1779
5591	rú	174.0365	169.9150	164.2592	158.8045	154.6704	150.9763	147.2686	143.3871	138.6858	133.9659	128.4217
5796	rö	103.5946	104.3598	105.0165	106.3065	110.0354	114.7238	119.1592	122.7599	125.7177	127.5593	130.3069
5807	rö	210.3324	200.7521	188.5765	179.0679	172.7375	168.3976	162.8347	156.4707	149.2341	141.2491	135.6740
5868	rò	95.6464	95.4354	95.1566	94.9286	94.6315	94.0123	93.5932	92.5909	91.2964	89.4979	87.4246
5904	ró	114.6160	118.7701	123.2932	127.7274	132.0720	135.7715	139.6171	143.2283	146.9604	150.5714	153.8145
6084	rì	89.8278	91.9655	92.9981	93.7721	95.1567	96.5677	98.0265	99.5412	100.3989	100.1362	99.9257
6131	rí	197.3165	191.9624	185.2218	178.2485	173.1483	168.1049	163.3125	158.6087	153.5073	149.9926	147.7551
6336	rê	94.7908	95.0096	96.0253	98.0619	99.3768	100.7763	100.7477	100.6010	99.7854	98.0745	95.2895
6347	rê	151.0071	143.3940	135.3423	129.5613	122.8102	115.8337	108.8712	101.1426	95.3914	87.5491	77.2093
6383	rè	174.1443	167.9768	161.0590	154.4093	147.0929	140.3120	132.8154	125.2644	116.8095	108.9242	100.5812
6444	re	102.5300	104.2159	106.7237	109.9657	113.1108	115.5739	117.1953	118.5326	120.2658	121.7172	121.9862
6588	rán	103.6592	106.7463	110.5681	113.7003	117.0128	120.7428	124.1899	127.5918	130.5301	133.2128	136.3235
6948	pu	119.6669	120.1871	120.2812	121.5033	122.4372	122.4504	122.6654	123.0807	123.6869	124.4807	125.2517
6959	pu	174.5557	165.9387	154.5329	145.9635	136.5809	126.7691	119.3324	111.9434	105.1094	99.2102	93.0700
7020	pön	136.5414	137.5673	139.2344	141.6772	144.9693	149.7539	153.9433	158.1365	162.4653	168.5872	175.9485
7031	pön	195.4338	187.0324	177.4424	169.0887	162.0680	156.1590	149.6166	143.1699	138.2129	133.9485	128.8799
7128	pö	130.2202	130.9704	133.7646	137.5568	140.5442	142.7349	144.6516	145.5691	148.1043	168.2457	183.9236
7139	pö	193.2768	184.4333	176.1794	169.7370	162.3425	154.3446	146.8027	139.3522	133.9666	136.5816	138.3581
7175	pô	171.1309	163.7154	154.4687	145.8284	137.4860	129.5704	122.1538	114.9013	107.6883	100.3744	96.0860
7211	pò	142.3099	133.5723	123.6602	114.0543	103.7964	95.0780	85.1727				
7272	po	120.3865	120.5121	121.4541	122.6869	123.4825	124.0364	125.1013	126.0461	126.8434	127.4974	
7344	pín	139.5456	141.5284	144.0085	147.6849	151.3667	156.4148	160.5640	163.7198	167.8586	175.3529	182.5421
7427	pì	143.9262	136.7140	124.7867	114.3435	103.5815	93.2548	84.2498	78.1413			
7452	pí	143.1520	147.6499	153.9392	156.9582	160.0166	167.6540	177.4438	186.4619	194.3550	200.9262	206.7099

7740	pe	122.1183	122.8530	124.0040	124.6958	125.5146	127.2916	128.9886	129.7889	131.5770	132.0410	132.3486
7992	pá	125.1481	126.1960	126.4430	128.1137	130.2569	134.7177	148.6715	155.8587	163.7572	170.5487	176.0600
8172	õ	102.5543	102.6171	101.9878	100.9551	99.6758	98.1604	96.5272	95.1005			
8208	ö	127.4035	129.7184	132.7403	135.9222	139.8575	143.2412	147.0108	150.3039	153.4383	158.1206	165.2854
8316	ó	131.0290	133.8359	137.6494	141.1306	143.6090	146.0720	148.7440	151.4880	154.9050	159.0928	164.7331
8568	nu	116.0371	120.4741	122.9006	125.3728	127.9535	131.1776	134.2913	135.2789	136.2629	138.7426	140.9736
8759	nö	182.7529	179.5636	176.3399	173.9731	168.3708	161.3567	156.1158	150.6430	145.0609	140.0493	135.0702
8820	nò	103.0723	103.9102	105.3424	108.3500	109.5654	109.8184	110.3222	110.2477	109.6510	108.1409	106.2129
9072	ní	124.9842	129.3413	134.6490	141.6548	146.7214	150.5791	155.5651	161.8433	167.7445	173.3981	181.3544
9227	në	217.1133	214.1167	206.6342	198.8371	192.2712	186.3646	178.4348	173.4134	166.1357	161.0506	154.8674
9371	ne	201.4201	194.5945	186.2822	177.7789	170.2256	161.3443	148.9259	139.4763	130.0613	124.0081	117.8293
9587	nà	190.1052	181.1539	171.0764	161.1857	152.9200	145.1191	135.4430	124.8803	114.4736	104.3692	94.0359
9659	na	200.2234	200.2172	189.4895	179.9458	170.6450	160.0780	151.8110	142.6642	133.9364	125.5463	118.0113
9720	n	123.7862	125.2350	126.8714	130.3454	132.7463	133.8631	134.9418	136.5609	137.7041	138.8782	140.3006
10152	mô	107.1538	109.2042	109.0776	107.1925	106.7948	107.2227	107.5699	107.9243	107.9318	107.6878	107.7892

Table 3. Extracted Intensity Features of Certain Syllables in Decibels

Id	S	I0(dB)	I1(dB)	I2(dB)	I3(dB)	I4(dB)	I5(dB)	I6(dB)	I7(dB)	I8(dB)	I9(dB)	I10(dB)
36	Yún	64.8758	64.8758	65.7774	66.7776	67.5933	68.2621	68.7898	69.2579	70.0243	71.1411	71.9663
47	Yún	65.5732	65.5732	66.3457	67.2626	67.9376	68.5739	69.1156	69.7630	70.6255	71.3720	71.9675
72	Yun	64.7718	64.7718	65.6936	66.6388	67.3154	67.8245	68.2929	68.7917	69.6456	70.6699	71.4188
108	yù	57.0904	57.0904	57.6380	58.8567	59.8386	60.3156	60.5941	60.9082	61.3838	61.7315	61.8655
119	yù	57.5987	57.5987	58.1732	59.5016	60.3884	60.6448	60.5966	60.4835	60.3088	60.2411	60.2011
155	yú	62.9795	62.9795	63.2426	63.3546	63.5281	63.8271	64.1828	64.5851	65.0217	65.6276	66.3441
828	yë	55.4108	55.4108	56.7362	57.1450	57.2538	57.3497	57.7792	59.1060	60.9664	62.9025	64.7454
972	ye	53.3657	53.3657	56.1304	57.8343	58.6381	59.1774	59.6484	60.0977	60.4730	60.8231	61.3204
1008	yçn	58.5995	58.5995	60.8219	63.5982	65.3735	66.6225	67.5551	68.2897	69.1501	69.9781	70.5797
1019	yçn	58.4976	58.4976	61.0766	64.2761	66.5383	67.7873	68.4374	68.9017	69.5557	70.1519	70.7019
2016	wí	66.2478	66.4523	68.4466	69.9957	70.7287	70.7718	70.4047	69.8431	69.2217	68.6014	67.5978
2027	wí	67.3716	67.4916	68.7091	70.0101	70.9468	71.0583	70.5617	69.6764	68.5997	67.5623	66.2718
2171	wë	60.4825	60.4825	61.7781	63.2475	64.7413	67.0802	70.0253	72.2494	73.0675	72.8114	71.9343
2232	wè	56.8570	56.8570	59.6401	62.4275	65.3647	67.9086	69.1272	69.7645	70.3422	70.6099	70.4999
2279	wé	60.6414	60.8855	63.4791	66.0773	67.6182	68.7196	70.0562	71.4034	72.9898	73.9584	73.8079
2315	we	69.6504	69.6504	70.9662	72.5125	73.8599	74.5349	74.9419	74.9839	74.6440	74.0324	73.2750
2376	wç	66.9189	66.9189	69.1911	70.7191	71.6599	72.2191	72.2420	72.0094	71.6949	71.2432	70.5502
2772	ù	65.3060	65.3060	65.5252	65.7338	65.8644	65.9367	65.8030	65.3791	64.8126	64.1643	63.4731
2808	ú	69.4604	69.4604	69.1059	69.1405	69.2024	69.0546	68.7879	68.5706	68.3573	67.8766	67.0544
2819	ú	70.5813	70.5813	69.9593	69.7344	69.6586	69.4883	69.1283	68.6987	68.0984	67.1576	65.8920
3024	tu	69.9946	69.9946	69.8191	70.3103	70.6365	70.8420	70.8771	70.5919	70.0060	68.8138	66.7449
3215	tö	68.1067	68.1067	67.6802	67.0087	66.4276	65.7070	64.8395	63.8919	62.8062	61.4779	59.5898
3240	tô	66.4072	66.4072	66.5321	66.7743	66.8917	66.7965	66.5623	66.2055	65.5343	64.2328	62.1547
3312	tó	74.0221	74.0221	74.1072	73.9892	73.6406	73.0278	72.0871	70.5464	68.2204	65.6711	65.1493
3575	ti	70.6521	70.6521	70.9681	71.2626	71.3690	71.4830	71.5028	71.4342	71.2254	70.5745	69.3257
3827	te	72.4462	72.4462	72.2117	71.9373	71.5566	70.9713	70.0548	68.9367	67.5622	65.7075	62.9565
4043	tà	72.1102	72.1102	71.3783	70.6725	69.6582	68.1484	66.5040	59.6662			
4068	tá	72.1033	72.1033	71.6538	71.2863	71.1214	71.0622	70.9071	70.6243	70.1656	69.4059	68.3312
4079	tá	73.0259	73.0259	72.4712	71.9041	71.5858	71.3888	70.9904	70.2825	69.2802	67.9712	66.6200
4187	sún	72.8295	72.8295	72.7769	72.8182	72.8585	72.7247	72.1778	71.3450	70.2721	68.9891	67.7936
4212	sun	73.3995	73.3995	73.4445	73.5851	73.7523	73.8383	73.7899	73.6177	73.3430	72.9323	72.3322
4223	sun	74.1242	74.1242	74.0315	73.9850	74.0670	74.1163	73.9646	73.4881	72.6853	71.7493	70.7835
4835	sí	72.3789	72.3789	72.3409	72.3738	72.3611	72.0965	71.2449	69.6598	67.5423	66.0996	66.0882
5256	sán	71.0546	71.0546	71.0774	71.0492	70.9961	70.9219	70.6919	70.0961	68.9707	67.3841	65.5560
5267	sán	71.5250	71.5250	71.4609	71.3824	71.2769	70.9694	70.3245	69.2732	67.6050	65.6634	63.5221
5292	san	69.3489	69.3489	69.5551	69.7219	69.7885	69.7355	69.6922	69.6495	69.4748	68.9801	68.0778
5483	rún	69.8888	69.8888	70.6152	71.2273	71.5509	71.8249	72.0877	72.1778	72.0220	71.5496	70.7706
5591	rú	69.4111	69.4928	70.2141	70.6695	71.0112	71.2041	71.3142	71.2225	70.8689	70.2970	69.3337
5796	rö	65.9102	66.1862	67.4233	68.4548	69.3765	70.0845	70.6786	71.0685	71.3475	71.5916	71.5845
5807	rö	67.6518	67.8805	68.9741	69.9381	70.7605	71.2923	71.5964	71.6438	71.6637	71.7840	71.6012
5868	rò	67.8667	67.8667	68.7288	69.4241	69.7647	69.8135	69.6261	69.1636	68.5152	67.6571	66.5097
5904	ró	69.0036	69.0036	69.5445	70.4098	71.0171	71.4933	71.8179	71.9343	71.9262	71.8161	71.5169

6084	rì	65.4920	65.4920	66.3456	67.4362	68.3878	69.2066	69.9149	70.3941	70.7794	70.9595	70.7049
6131	rí	63.5512	63.5512	64.5365	65.4892	66.3540	67.1868	67.9900	68.4699	68.6763	68.3389	67.4754
6336	rê	63.9859	63.9859	65.3227	66.9859	68.4935	69.2959	69.3384	69.3502	69.3167	68.5172	66.8846
6347	rê	65.9556	65.9556	67.0644	68.4265	69.0848	69.5277	69.6919	69.1959	68.6685	68.0600	66.3967
6383	rè	67.1522	67.1522	68.2394	69.1107	69.7496	70.1082	70.2649	70.2544	70.0793	69.7266	69.1108
6444	re	65.8679	65.8679	66.6605	67.5559	68.1994	68.6823	68.9085	68.9288	68.8379	68.5985	68.2055
6588	rán	63.5597	63.5597	65.2559	66.4126	67.0933	67.5294	67.8055	67.9528	67.9866	68.0050	67.9437
6948	pu	66.2980	66.2980	66.2567	66.4278	66.9227	67.1344	66.9372	66.4839	65.4922	63.8971	61.9568
6959	pu	67.1171	67.1171	67.4826	67.1708	67.2332	67.2601	66.6631	65.8089	64.6889	62.9965	60.4148
7020	pön	68.1546	68.1546	68.1817	68.1111	67.8852	67.5441	66.9927	66.2014	65.1048	63.7920	62.2038
7031	pön	68.6717	68.6717	68.6592	68.5067	68.2133	67.7132	66.7741	65.4726	63.8557	61.8729	59.8717
7128	pö	65.5983	65.5983	66.6513	67.6030	68.1031	68.0593	67.5235	66.0145	63.7548	62.1635	62.0135
7139	pö	65.9809	65.9809	67.1750	68.0430	68.4178	68.3632	67.7633	66.1112	63.3152	60.5972	59.8722
7175	pô	69.9095	69.9095	70.0355	70.2346	70.4997	70.7197	70.4564	69.7462	68.9580	67.9578	66.2535
7211	pò	72.1619	72.1619	71.9208	71.0511	70.0060	68.9749	67.8043				
7272	po	69.8964	69.8964	69.9725	69.9748	69.7970	69.4157	68.8930	68.0391	67.2118	67.2118	
7344	pín	69.6533	69.6533	69.8540	70.0773	70.1312	69.9348	69.4369	68.5132	67.1911	65.7252	64.3901
7427	pì	71.7851	71.7851	71.5607	70.9341	69.8616	68.2457	66.1804	64.1855			
7452	pí	71.5142	71.5142	71.1367	70.3484	69.4987	68.7181	68.3338	68.4575	68.6986	68.2807	67.4140
7740	pe	71.9921	71.9921	71.9101	71.6631	71.3612	71.0547	70.6887	70.1842	69.6297	69.0064	68.0656
7992	pá	67.1554	67.1554	66.6923	65.8348	64.7684	63.4574	62.3107	61.6370	61.4320	61.1886	58.8585
8172	õ	65.8035	65.8035	65.8275	65.5996	65.1643	64.5976	63.9483	63.9483			
8208	ö	67.5748	67.5748	67.5441	67.4082	67.2674	67.1075	66.8804	66.5063	65.8144	64.9516	64.3012
8316	ó	71.9668	71.9668	72.3293	72.5266	72.5673	72.4637	72.1208	71.5316	70.7358	69.8660	69.3016
8568	nu	65.9230	66.1334	67.0756	67.7844	68.3939	68.9164	69.6141	70.4901	70.9643	71.0570	71.0481
8759	nö	64.5587	64.5587	65.0211	65.5692	66.6588	68.1045	68.7791	69.0113	69.0277	68.8341	68.5828
8820	nò	69.3339	69.3339	70.1204	71.0390	71.7462	72.1483	72.3603	72.4674	72.5999	72.4758	72.0603
9072	ní	66.0352	66.2922	67.5040	68.6034	69.7597	70.4559	70.7360	70.7707	70.6068	70.1952	69.5920
9227	në	67.3580	67.3580	68.1681	69.0053	69.7971	70.9881	72.4532	73.0856	72.9774	72.5072	71.8388
9371	ne	67.0872	67.0872	67.3414	67.5814	67.9202	68.8308	70.1420	70.8649	70.9370	70.6714	70.2189
9587	nà	66.0359	66.0359	66.9357	67.5799	68.1733	68.8598	69.8327	70.4715	70.2825	69.2997	68.1224
9659	na	68.9785	68.9785	70.2739	71.0969	71.2833	71.1675	70.9362	70.6058	70.0747	69.2323	68.2690
9720	n	64.8434	64.8434	65.1662	65.7849	66.3928	66.6221	66.6628	66.6083	66.5142	66.4261	66.3211
10152	mô	64.1690	64.1690	66.5050	67.9343	68.2302	68.2106	68.2462	68.2492	68.1524	67.9502	67.4341

Table 4. Extracted Duration Features of Certain Syllables in Seconds

Id	S	T0(dB)	T1(dB)	T2(dB)	T3(dB)	T4(dB)	T5(dB)	T6(dB)	T7(dB)	T8(dB)	T9(dB)	T10(dB)
36	yún	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
47	yún	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
72	yun	0.0227	0.0327	0.0427	0.0527	0.0627	0.0727	0.0827	0.0927	0.1027	0.1127	0.1227
108	yù	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
119	yù	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
155	yú	0.0221	0.0321	0.0421	0.0521	0.0621	0.0721	0.0821	0.0921	0.1021	0.1121	0.1221
828	yë	0.0231	0.0331	0.0431	0.0531	0.0631	0.0731	0.0831	0.0931	0.1031	0.1131	0.1231
972	ye	0.0233	0.0333	0.0433	0.0533	0.0633	0.0733	0.0833	0.0933	0.1033	0.1133	0.1233
1008	yçn	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
1019	yçn	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
2016	wí	0.0233	0.0333	0.0433	0.0533	0.0633	0.0733	0.0833	0.0933	0.1033	0.1133	0.1233
2027	wí	0.0233	0.0333	0.0433	0.0533	0.0633	0.0733	0.0833	0.0933	0.1033	0.1133	0.1233
2171	wë	0.0207	0.0307	0.0407	0.0507	0.0607	0.0707	0.0807	0.0907	0.1007	0.1107	0.1207
2232	wè	0.0223	0.0323	0.0423	0.0523	0.0623	0.0723	0.0823	0.0923	0.1023	0.1123	0.1223
2279	wé	0.0231	0.0331	0.0431	0.0531	0.0631	0.0731	0.0831	0.0931	0.1031	0.1131	0.1231
2315	we	0.0234	0.0334	0.0434	0.0534	0.0634	0.0734	0.0834	0.0934	0.1034	0.1134	0.1234
2376	wç	0.0218	0.0318	0.0418	0.0518	0.0618	0.0718	0.0818	0.0918	0.1018	0.1118	0.1218
2772	ù	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900	0.1000	0.1100	0.1200

2808	ú	0.0249	0.0349	0.0449	0.0549	0.0649	0.0749	0.0849	0.0949	0.1049	0.1149	0.1249
2819	ú	0.0249	0.0349	0.0449	0.0549	0.0649	0.0749	0.0849	0.0949	0.1049	0.1149	0.1249
3024	tu	0.0248	0.0348	0.0448	0.0548	0.0648	0.0748	0.0848	0.0948	0.1048	0.1148	0.1248
3215	tö	0.0207	0.0307	0.0407	0.0507	0.0607	0.0707	0.0807	0.0907	0.1007	0.1107	0.1207
3240	tô	0.0235	0.0335	0.0435	0.0535	0.0635	0.0735	0.0835	0.0935	0.1035	0.1135	0.1235
3312	tó	0.0227	0.0327	0.0427	0.0527	0.0627	0.0727	0.0827	0.0927	0.1027	0.1127	0.1227
3575	ti	0.0230	0.0330	0.0430	0.0530	0.0630	0.0730	0.0830	0.0930	0.1030	0.1130	0.1230
3827	te	0.0226	0.0326	0.0426	0.0526	0.0626	0.0726	0.0826	0.0926	0.1026	0.1126	0.1226
4043	tà	0.0239	0.0339	0.0439	0.0539	0.0639	0.0739	0.0839	0.1339			
4068	tá	0.0220	0.0320	0.0420	0.0520	0.0620	0.0720	0.0820	0.0920	0.1020	0.1120	0.1220
4079	tá	0.0220	0.0320	0.0420	0.0520	0.0620	0.0720	0.0820	0.0920	0.1020	0.1120	0.1220
4187	sún	0.0236	0.0336	0.0436	0.0536	0.0636	0.0736	0.0836	0.0936	0.1036	0.1136	0.1236
4212	sun	0.0204	0.0304	0.0404	0.0504	0.0604	0.0704	0.0804	0.0904	0.1004	0.1104	0.1204
4223	sun	0.0204	0.0304	0.0404	0.0504	0.0604	0.0704	0.0804	0.0904	0.1004	0.1104	0.1204
4835	sí	0.0239	0.0339	0.0439	0.0539	0.0639	0.0739	0.0839	0.0939	0.1039	0.1139	0.1239
5256	sán	0.0217	0.0317	0.0417	0.0517	0.0617	0.0717	0.0817	0.0917	0.1017	0.1117	0.1217
5267	sán	0.0217	0.0317	0.0417	0.0517	0.0617	0.0717	0.0817	0.0917	0.1017	0.1117	0.1217
5292	san	0.0230	0.0330	0.0430	0.0530	0.0630	0.0730	0.0830	0.0930	0.1030	0.1130	0.1230
5483	rún	0.0210	0.0310	0.0410	0.0510	0.0610	0.0710	0.0810	0.0910	0.1010	0.1110	0.1210
5591	Rú	0.0250	0.0350	0.0450	0.0550	0.0650	0.0750	0.0850	0.0950	0.1050	0.1150	0.1250
5796	Rö	0.0242	0.0342	0.0442	0.0542	0.0642	0.0742	0.0842	0.0942	0.1042	0.1142	0.1242
5807	Rö	0.0242	0.0342	0.0442	0.0542	0.0642	0.0742	0.0842	0.0942	0.1042	0.1142	0.1242
5868	Rò	0.0243	0.0343	0.0443	0.0543	0.0643	0.0743	0.0843	0.0943	0.1043	0.1143	0.1243
5904	Ró	0.0210	0.0310	0.0410	0.0510	0.0610	0.0710	0.0810	0.0910	0.1010	0.1110	0.1210
6084	Rì	0.0203	0.0303	0.0403	0.0503	0.0603	0.0703	0.0803	0.0903	0.1003	0.1103	0.1203
6131	Rí	0.0220	0.0320	0.0420	0.0520	0.0620	0.0720	0.0820	0.0920	0.1020	0.1120	0.1220
6336	Rê	0.0228	0.0328	0.0428	0.0528	0.0628	0.0728	0.0828	0.0928	0.1028	0.1128	0.1228
6347	Rê	0.0228	0.0328	0.0428	0.0528	0.0628	0.0728	0.0828	0.0928	0.1028	0.1128	0.1228
6383	Rè	0.0245	0.0345	0.0445	0.0545	0.0645	0.0745	0.0845	0.0945	0.1045	0.1145	0.1245
6444	Re	0.0201	0.0301	0.0401	0.0501	0.0601	0.0701	0.0801	0.0901	0.1001	0.1101	0.1201
6588	rán	0.0226	0.0326	0.0426	0.0526	0.0626	0.0726	0.0826	0.0926	0.1026	0.1126	0.1226
6948	Pu	0.0239	0.0339	0.0439	0.0539	0.0639	0.0739	0.0839	0.0939	0.1039	0.1139	0.1239
6959	Pu	0.0239	0.0339	0.0439	0.0539	0.0639	0.0739	0.0839	0.0939	0.1039	0.1139	0.1239
7020	pön	0.0201	0.0301	0.0401	0.0501	0.0601	0.0701	0.0801	0.0901	0.1001	0.1101	0.1201
7031	pön	0.0201	0.0301	0.0401	0.0501	0.0601	0.0701	0.0801	0.0901	0.1001	0.1101	0.1201
7128	Pö	0.0249	0.0349	0.0449	0.0549	0.0649	0.0749	0.0849	0.0949	0.1049	0.1149	0.1249
7139	Pö	0.0249	0.0349	0.0449	0.0549	0.0649	0.0749	0.0849	0.0949	0.1049	0.1149	0.1249
7175	Pô	0.0203	0.0303	0.0403	0.0503	0.0603	0.0703	0.0803	0.0903	0.1003	0.1103	0.1203
7211	Pò	0.0217	0.0317	0.0417	0.0517	0.0617	0.0717	0.0817				
7272	Po	0.0228	0.0328	0.0428	0.0528	0.0628	0.0728	0.0828	0.0928	0.1028	0.1128	
7344	pín	0.0201	0.0301	0.0401	0.0501	0.0601	0.0701	0.0801	0.0901	0.1001	0.1101	0.1201
7427	Pì	0.0204	0.0304	0.0404	0.0504	0.0604	0.0704	0.0804	0.0904			
7452	Pí	0.0242	0.0342	0.0442	0.0542	0.0642	0.0742	0.0842	0.0942	0.1042	0.1142	0.1242
7740	Pe	0.0249	0.0349	0.0449	0.0549	0.0649	0.0749	0.0849	0.0949	0.1049	0.1149	0.1249
7992	Pá	0.0227	0.0327	0.0427	0.0527	0.0627	0.0727	0.0827	0.0927	0.1027	0.1127	0.1227
8172	Õ	0.0246	0.0346	0.0446	0.0546	0.0646	0.0746	0.0846	0.0946			
8208	Ö	0.0202	0.0302	0.0402	0.0502	0.0602	0.0702	0.0802	0.0902	0.1002	0.1102	0.1202
8316	Ó	0.0223	0.0323	0.0423	0.0523	0.0623	0.0723	0.0823	0.0923	0.1023	0.1123	0.1223

8568	Nu	0.0246	0.0346	0.0446	0.0546	0.0646	0.0746	0.0846	0.0946	0.1046	0.1146	0.1246
8759	Nö	0.0248	0.0348	0.0448	0.0548	0.0648	0.0748	0.0848	0.0948	0.1048	0.1148	0.1248
8820	Nò	0.0206	0.0306	0.0406	0.0506	0.0606	0.0706	0.0806	0.0906	0.1006	0.1106	0.1206
9072	Ní	0.0249	0.0349	0.0449	0.0549	0.0649	0.0749	0.0849	0.0949	0.1049	0.1149	0.1249
9227	Në	0.0233	0.0333	0.0433	0.0533	0.0633	0.0733	0.0833	0.0933	0.1033	0.1133	0.1233
9371	Ne	0.0211	0.0311	0.0411	0.0511	0.0611	0.0711	0.0811	0.0911	0.1011	0.1111	0.1211
9587	Nà	0.0245	0.0345	0.0445	0.0545	0.0645	0.0745	0.0845	0.0945	0.1045	0.1145	0.1245
9659	Na	0.0235	0.0335	0.0435	0.0535	0.0635	0.0735	0.0835	0.0935	0.1035	0.1135	0.1235
9720	N	0.0213	0.0313	0.0413	0.0513	0.0613	0.0713	0.0813	0.0913	0.1013	0.1113	0.1213
10152	mô	0.0235	0.0335	0.0435	0.0535	0.0635	0.0735	0.0835	0.0935	0.1035	0.1135	0.1235

Table 5. Mean Values for Extracted Features of Syllables

Id	S	NF	SP (s)	D (s)	MPC (Hz)	MPP (Hz)	SDPC	SDPP	ETP (s)
36	yún	26	0.0000	0.2904	152.0796	152.0100	43.8074	27.4252	0.2904
47	yún	25	0.0000	0.2904	143.6029	144.8873	61.9963	48.7294	0.2904
72	yun	29	0.0000	0.3253	121.6787	122.0979	27.4586	10.6191	0.3253
108	yù	16	0.0000	0.1903	99.2542	99.4725	25.9855	2.0512	0.1903
119	yù	11	0.0000	0.1903	99.3778	107.1205	42.4009	22.1473	0.1903
155	yú	27	0.0000	0.3142	140.2745	141.7953	55.6729	41.4099	0.3142
828	yë	25	0.0000	0.2861	132.7996	132.8151	40.5266	25.0702	0.2861
972	ye	33	0.0000	0.3667	110.6201	110.6526	26.3368	13.0177	0.3667
1008	yçn	30	0.0000	0.3304	110.5507	110.7147	24.1665	9.8441	0.3304
1019	yçn	23	0.0000	0.3304	130.5383	137.0289	55.5117	42.7529	0.3304
2016	wí	18	0.0000	0.2165	167.1898	167.1921	56.4727	31.8892	0.2165
2027	wí	18	0.0000	0.2165	149.4088	149.9393	49.0773	27.6795	0.2165
2171	wë	21	0.0000	0.2515	132.3561	134.2253	45.6280	30.1191	0.2515
2232	wè	20	0.0000	0.2346	97.2361	98.6374	22.0786	5.8115	0.2346
2279	wé	22	0.0000	0.2563	127.3490	127.5030	46.5140	32.0518	0.2563
2315	we	16	0.0000	0.2368	115.6685	123.8763	48.8091	32.1188	0.2368
2376	wç	20	0.0000	0.2337	115.4044	115.4662	29.8982	7.6477	0.2337
2772	ù	18	0.0000	0.2101	93.2084	93.7755	22.7850	5.0952	0.2101
2808	ú	19	0.0000	0.2297	182.1944	181.1096	59.3558	26.3641	0.2297
2819	ú	19	0.0000	0.2297	167.8146	166.6569	65.2970	42.0648	0.2297
3024	tu	15	0.0000	0.1896	119.5114	119.7731	34.9410	2.0615	0.1896
3215	tö	15	0.0000	0.1814	142.9495	141.5081	54.5402	31.5861	0.1814
3240	tô	17	0.0000	0.2070	113.7873	112.7797	33.7060	4.9464	0.2070
3312	tó	13	0.0000	0.1654	166.3811	164.9713	55.7977	16.7820	0.1654
3575	ti	17	0.0000	0.2160	122.3120	123.4541	53.0796	36.3502	0.2160
3827	te	15	0.0000	0.1851	126.2890	124.3567	54.2306	33.6969	0.1851
4043	tà	8	0.0000	0.1577	100.2770	106.1232	44.2988	23.4288	0.1577
4068	tá	16	0.0000	0.1939	153.5096	152.9872	45.4450	13.1749	0.1939
4079	tá	16	0.0000	0.1939	143.2690	141.8769	56.5377	34.2700	0.1939
4187	sún	15	0.0000	0.1873	174.8550	173.3948	64.3995	32.7395	0.1873
4212	sun	18	0.0000	0.2109	136.4367	136.0641	36.3782	8.0152	0.2109
4223	sun	17	0.0000	0.2109	131.3698	132.1017	57.8670	39.5015	0.2109
4835	sí	15	0.0000	0.1877	169.8467	169.4228	59.5798	29.7671	0.1877
5256	sán	14	0.0000	0.1734	154.5537	153.4527	50.0988	16.3825	0.1734
5267	sán	14	0.0000	0.1734	141.0734	139.9134	51.4278	26.1309	0.1734
5292	san	19	0.0000	0.2260	108.8811	108.2577	30.2288	5.6276	0.2260
5483	rún	21	0.0000	0.2420	133.8847	133.3840	50.6419	35.3982	0.2420
5591	rú	18	0.0000	0.2199	135.2542	134.8845	46.3850	24.6042	0.2199
5796	rö	25	0.0000	0.2884	138.8990	138.6745	42.4067	24.6587	0.2884
5807	rö	23	0.0000	0.2884	131.3259	134.1779	56.4954	40.2940	0.2884
5868	rò	15	0.0000	0.1885	90.0719	90.4470	26.3164	5.1720	0.1885

5904	ró	20	0.0000	0.2320	150.6295	150.5199	43.9947	21.1128	0.2320
6084	rì	19	0.0000	0.2207	94.0946	94.9447	21.6892	4.0459	0.2207
6131	rí	24	0.0000	0.2740	135.2617	135.0051	51.6648	37.8659	0.2740
6336	rê	16	0.0000	0.2156	91.3141	93.6679	23.7614	7.6764	0.2156
6347	rê	11	0.0000	0.2156	102.7262	115.2829	45.5482	23.6329	0.2156
6383	rè	17	0.0000	0.2690	113.0376	120.0446	48.7815	32.9268	0.2690
6444	re	22	0.0000	0.2502	118.7393	119.1509	27.7281	7.6425	0.2502
6588	rán	23	0.0000	0.2652	138.9822	139.4407	39.2851	21.3245	0.2652
6948	pu	15	0.0000	0.1879	124.8141	124.1181	37.9155	4.0678	0.1879
6959	pu	15	0.0000	0.1879	118.7901	117.0187	51.8350	32.5769	0.1879
7020	pön	15	0.0000	0.1801	159.6729	159.7098	47.0880	17.7541	0.1801
7031	pön	15	0.0000	0.1801	146.6533	146.1834	51.4669	28.2446	0.1801
7128	pö	16	0.0000	0.1997	164.5208	163.4983	58.1787	30.0853	0.1997
7139	Pö	16	0.0000	0.1997	148.1998	147.1462	52.3768	24.0293	0.1997
7175	Pô	15	0.0000	0.1906	117.6896	118.7303	47.7518	30.5091	0.1906
7211	Pò	7	0.0000	0.1233	109.2265	113.9491	48.1976	20.7002	0.1233
7272	Po	10	0.0000	0.1356	123.8407	123.8047	41.5913	2.5611	0.1356
7344	Pín	14	0.0000	0.1702	164.3554	164.1827	50.3244	18.4331	0.1702
7427	Pì	8	0.0000	0.1308	105.2972	109.8747	46.5141	24.2064	0.1308
7452	Pí	15	0.0000	0.1884	182.3439	183.3684	58.6645	26.7084	0.1884
7740	Pe	17	0.0000	0.2099	132.2737	132.1263	38.3542	7.5802	0.2099
7992	Pá	13	0.0000	0.1655	150.5219	149.8611	52.0357	22.5772	0.1655
8172	Õ	8	0.0000	0.1191	99.4116	99.6973	36.6573	2.8550	0.1191
8208	Ö	19	0.0000	0.2204	163.6149	163.6402	48.6906	26.2607	0.2204
8316	Ó	18	0.0000	0.2246	129.1719	131.5997	36.4759	25.8920	0.2246
8568	Nu	25	0.0000	0.2891	140.5224	141.6423	33.4255	12.1187	0.2891
8759	Nö	20	0.0000	0.2395	141.3034	140.9247	47.2002	26.1045	0.2395
8820	Nò	18	0.0000	0.2112	102.6451	103.7473	24.5550	6.3203	0.2112
9072	Ní	17	0.0000	0.2099	171.1889	170.6888	59.9074	31.1589	0.2099
9227	Në	21	0.0000	0.2466	158.6528	158.0243	56.0524	35.1576	0.2466
9371	Ne	18	0.0000	0.2522	125.0962	133.6226	55.4558	40.6220	0.2522
9587	Nà	13	0.0000	0.2289	118.2873	133.3755	56.5511	37.4787	0.2289
9659	Na	16	0.0000	0.2370	129.6426	140.4555	57.5689	40.2845	0.2370
9720	N	27	0.0000	0.3026	140.1785	140.1660	31.7942	8.1726	0.3026
10152	Mô	19	0.0000	0.2271	108.4655	107.6228	30.4556	3.4576	0.2271

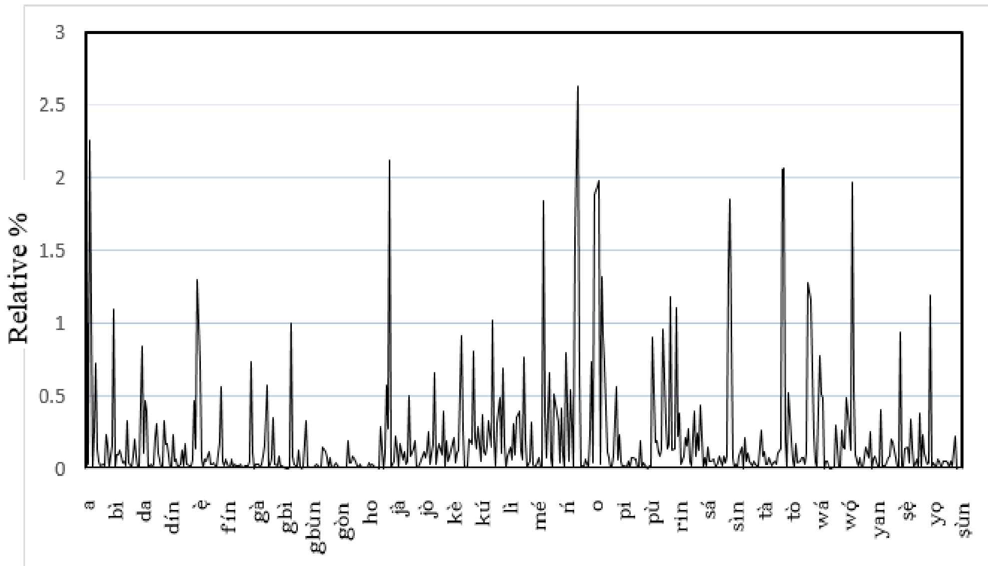

Figure 12: Relative % of Corporal Syllables in Selection (Alphabetic order)

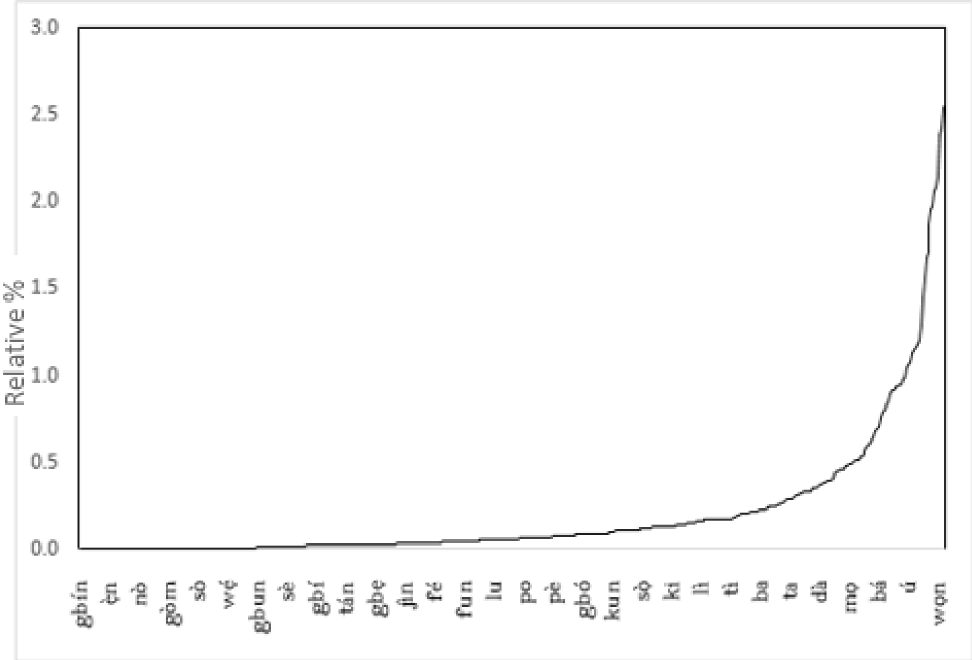

Figure 13: Relative % of Corporal Syllables in Selection (Ascending order)

6. SYSTEM EVALUATION

6.1 Mean Opinion Score (MOS)

The developed corpus was used to run a Standard Yoruba Text-to-Speech system. A mean opinion score was computed for the perceived overall quality of the systemstimuli. This measured the speech naturalness component of prosody. The MOS was estimated as an arithmetic mean for all individual.

30 Selected listeners were played some speech generated with the database and asked to rate what they perceived. The Listeners were told to indicate their assessments on a scale of 1 to 5. The grading is shown on Table 6.

Table 6. Rating and Grading Used for MOSTest

Rating	Grade
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

6.2 Semantically Unpredictable Sentences (SUS) Test

Intelligibility test of the corpus was carried out by usingthe semantically unpredictable sentences to estimate the quality of speech produced by the Standard Yoruba Text-to-Speech system using the corpus. The gave a measure of word recognition. In semantically unpredictable sentences (SUS) test, a subject in a quiet room listens to a series of 90 short semantically unpredictable (nonsense) sentences over loudspeakers. Some of the sentences were difficult to understand. The subject heard each sentence once, then he rated the intelligibility of the sentence on a scale from 1 -7.

Naturalness and Intelligibility tests were performed to estimate the quality of speech produced bythe Standard Yoruba Text-to-Speech system using the corpus. mean opinion score and semantically unpredictable sentences (SUS) test conducted with 30 subjects who understand and use SY regularly. The following results were obtained from 30 subjects. Table 7 shows the Rating and Grading Used for SUS Test while Table 8 presents the Result of SUS Test.

Table 7. Rating and Grading Used for SUSTest

Rating	Grade
7	Ideal
6	Excellent
5	Good
4	Fair
3	Poor
2	Bad
1	Extremely Bad

Table 8. Result of SUS Test

SN	SUBJECT	MOS NATURALNESS (1-5)	MOS NATURALNESS (%)	SUS INTELLIGIBILITY (1-7)	INTELLIGIBILITY (%) SUS
1	Subj1	3	60	6	85.7
2	Subj2	4	80	6	85.7
3	Subj3	2	40	7	100
4	Subj4	4	80	6	85.7
5	Subj5	4	80	7	100
6	Subj6	4	80	6	85.7
7	Subj7	4	80	7	100
8	Subj8	4	80	6	85.7
9	Subj9	4	80	6	85.7
10	Subj10	3	60	7	100
11	Subj11	4	80	7	100
12	Subj12	3	60	7	100
13	Subj13	4	80	6	85.7
14	Subj14	3	60	6	85.7
15	Subj15	4	80	7	100
16	Subj16	4	80	7	100
17	Subj17	2	40	7	100
18	Subj18	4	80	6	85.7
19	Subj19	3	60	6	85.7
20	Subj20	4	80	6	85.7
21	Subj21	4	80	7	100
22	Subj22	4	80	6	85.7
23	Subj23	4	80	7	100
24	Subj24	3	60	7	100
25	Subj25	4	80	6	85.7
26	Subj26	4	80	7	100
27	Subj27	3	60	7	100
28	Subj28	4	80	7	100
29	Subj29	4	80	7	100
30	Subj30	4	80	6	85.7
Arithmetic Mean		3.6	72.7	6.5	93.3

Intelligibility test is carried out with the modified rhyme test or semantically unpredictable sentences. These give a measure of word recognition. In semantically unpredictable sentences (SUS) test, a subject in a quiet room listens to a series of 90 short semantically unpredictable (nonsense) sentences over loudspeakers. Some of the sentences may be difficult to understand. The subject hears each sentence once, then he rates the intelligibility of the sentence on a scale from 1 -7.

Naturalness and Intelligibility tests were performed to estimate the perceptive quality of ProTToSSYL speech by way of mean opinion score and semantically unpredictable sentences (SUS) test conducted with 30 subjects who understand and use SY regularly. The following results were obtained from 30 subjects.

These tests show that the corpus produced a synthetic speech of MOS = 3.6 (72.7%) naturalness and SUS Score = 6.5 (93.3%) intelligibility.

7. CONCLUSION

This research developed a prosodic read corpus for the Standard Yoruba languageby extracting textual inputs from four sources: two Standard Yoruba (SY) fiction, an SY grammar textbook and an SY Online Scripture. The author applied a hybrid of Falaschi scheme and the add-on procedure of Radová and Vopálka to obtain phonetically balanced text bag containing 7376 phrases and sentences. The selected text was carefully read by an expert and recorded in a suitable environment and saved as wave files. The wave files were annotated with Praat. The relational database developed hosts the corpus metadata.

The corpus performed impressively when tested with a Standard Yoruba TTS.72.7%naturalness and 93.3% intelligibility were recorded.

Yoruba Language is tonal, exhibiting a tonality of 3. The tonality is not fixed. By implication, the syllables havesurface tones which are usually pitched at relatively dissimilar perceptual frequency levels. The syllable is a natural and convenient base unit for speech in Yoruba Language. Thus, the corpus is syllable based.

The result shows a naturalness of 72.7% when the corpus was used with a TTS. It is believed that limitation could have arose from perhaps miss syllabification of words in the corpus, and or absence of prosodic representation in written text. Error of syllabification could have arisen from the factthat the syllabification algorithm used a Declarative Rules-Based Approach (DRSA) for the syllabification of SY words.The DRSA would fail to completely disambiguate all the CV.N and DV.N syllabic structures. Consequently, it would misidentify some of the /n/ variant of the CV.N and DV.N syllable structures. Hence there is need to improve upon the syllabification algorithm used.

Moreover, if larger number of sentences were used, there could have been a wider coverage of the prosodic space than was is observed with 7376 phrases and sentences.

This study has been able to contribute a speech corpus to the Standard Yoruba NLP research.

8. REFERENCES

[1] Atoye R. O. 1999. “Native -Speaker Perception of Intonation in Yoruba Zero- Particle Interrogative Clauses.” Papers in English and Linguistics 4:15 - 23, 1999.

[2] Ajolore O. 1974. Learning to Use Yorùbá focus sentence in a multilingual setting. Ph.D. Thesis, University of

Ilinouis, USA, 1974.

[3] Adetugbo A. 1973. The Yoruba Language in Yoruba History. in Biobaku, S.O. (ed.). Sources of Yoruba History. pp. 176–204.

[4] Crowther S. 1843. Vocabulary of the Yorùbá Language… to which are prefixed the grammatical elements of the Yoruba language. Church Missionary Society, London.

[5] Fagborun J. G. 1994. The Yoruba Koiné – its History and Linguistic Innovations. LINCOM Linguistic Edition Vol. 6. München/Newcastle: LINCOM Europe.

[6] Akinlabi A. and Liberman M. 2000. The tonal phonology of Yoruba clitics. Rutgers Univ. & Univ. of Pennsylvania.

[7] Adewole L. O. 1987. The Yoruba Language: Published works and doctoral dissertations, 1843 - 1986. Hamburg: Helmut Buske.

[8] Taylor, P. 2009. Text-to-speech synthesis. Cambridge, UK: Cambridge University Press. pp. 3.

[9] Fajobi E. 2003. Why does Yoruba high tone fall where it does in 'Yorubalized' English words?" In Kingsley Sage (ed.). The 16th White House Papers: Graduate Research in Cognitive and Computing Sciences at Sussex. pp.34 - 36.

[10] Beckman, M.E., 1996. The parsing of prosody. Language and cognitive processes, 11(1-2), pp.17-68.

[11] Wennerstrom A. K.2001. The music of everyday speech: Prosody and discourse analysis. Oxford: Oxford University Press, 2001.

[12] Bird, S. 2011. Strategies for representing tone in African writing systems: a critical review. URL:http://cogprints.org/2174/00/wll2.pdf. (Access date: 7 July 2011).

[13] Oyetade B. A. 1987. Tone representation and Yoruba tone. Paper presented at UCL Dept. of Phonetics and Linguistics Postgraduate Seminar in March, 1987 and the Spring Meeting of the LAGB, Westfield College, London, April.

[14] Hu, C. 2003. Text statistics tool box for natural language processing. Technical report, The University of Georgia, May.

[15] Ngugi, K., Okelo-Odongo, W. and Wagacha, P.W., 2005. Swahili text-to-speech system. African Journal of Science and Technology, 6(1).

[16] Onayemi A. O. 2010, "Learn Yoruba", http://www.learnyoruba.com/ aboutus.htm. Retrieved on June 20, 2021

[17] Akinwonmi, A. E. and Alese, B. K. 2013. A prosodic text-to-speech system for the Yorùbá language. The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013). Technically Co-sponsored by IEEE UK/RI Computer Chapter. December 9-12, 2013, in London, UK. Pages 634 – 639 of the Conference Proceedings. ISBN 978-1-908320-16-2.

[18] Adedjouma Sèmiyou, A., Aoga, J.O. and Igue, M.A., 2012. Part-of-Speech tagging of Yorùbá Standard, Language of Niger-Congo family. Research Journal of Computer and Information Technology Sciences. Vol. 1 pages 2-5.

[19] Olaniyan, O. L. 2015. Development of a Yoruba Language Syllabificator. (Postgraduate diploma dissertation). The Federal University of Technology, Akure, Ondo State. Nigeria.

[20] Agrawal, S. S., Samudravijaya, K. and Arora, K. 2006. Recent advances of speech databases development activity for Indian languages, In Proc. ISCSLP 2006, Singapore, pp. 771-776.

[21] Kumolalo, F. O., Adagunodo, E. R., and Odejobi, O. A. 2010. Development of a Syllabicator for Yoruba Language. In Proceedings of OAU TekConf, September 5-8, 2010, pages 47–51, OAU, Ile-Ife, Nigeria.

[22] Broder, A.Z., Glassman, S.C., Manasse, M.S. and Zweig, G., 1997. Syntactic clustering of the web. Computer networks and ISDN systems, 29(8-13), pp.1157-1166.