EUROPEAN ACADEMIC RESEARCH
Vol. II, Issue 10/ January 2015

ISSN 2286-4822
www.euacademic.org

Impact Factor: 3.1 (UIF)
DRJI Value: 5.9 (B+)

A *Bisaya* Text-to-Speech (TTS) System Utilizing Rule-Based Algorithm and Concatenative Speech Synthesis

AILEN B. GARCIA
Information Technology Department
Occidental Mindoro State College
San Jose, Occidental Mindoro, Philippines
JAY NOEL N. ROJO
CONSORCIO S. NAMOCO, JR.
College of Industrial and Information Technology
Mindanao University of Science and Technology
Cagayan de Oro City, Philippines

Abstract:

Bisaya dialect is an Austronesian language spoken in the Philippines by about 20 million people. It is most widely spoken member of the Visayan language; a language closely related to Malaysian, Indonesian and the Filipino people. It is the second most spoken language in the Philippines.

This study developed a Text-to-speech (TTS) for Bisaya dialect by creating an algorithm based on the identified rules in Bisaya dialect syllabification. Bisaya word (phonemes) has morphophonemic processes which are composed of assimilation, deletion, alteration and metathesis. With these processes, eleven (11) rules for the syllabification process have been determined. These rules (algorithm) were the basis for text selection (as to the syllabication of the word) which were then used in concatenation or appending of the pre-recorded speech. Evaluation of the present Bisaya Text-to-speech system showed that it is more intelligible, natural and understandable as compared to the existing TextAloud TTS system which is designed for English language. This will be a good start for Bisaya Text-to-speech researches and may find possible market in mobile phone *services as application for the benefit of the visually impaired person and an eventful person.*

Key words: Text-to-speech, concatenative speech synthesis, rule-based, *Bisaya* dialect, syllabification, phonemes

Introduction

Text-to-speech (TTS) is a type of speech synthesis application that is used to create a spoken sound version of the text in a computer document (Text-to-Speech, 2001). It enables the reading of computer display information for the visually challenged person, or may simply be used to augment the reading of a text message. It is the process which artificially produces synthetic speech for various applications such as services over telephone, e-document, reading and speaking for handicapped people (vision-impaired person). Speech synthesis is a computer-generated simulation of human speech (Schroeder, 1993). The main techniques for speech synthesis design (Khalifa, 2011) are articulatory synthesis, formant synthesis, and concatenative synthesis. Articulatory synthesis attempts to model the human speech production system directly. On the other hand, formant synthesis models the pole of frequencies of speech signal or transfer function of vocal tract based on source-filter-model. In the case of concatenative synthesis, it uses different length pre-recorded samples derived from natural speech. The synthesis technologies may vary from simple speech concatenation to parametric synthesizers or full text-to-speech systems with unit selection synthesis. The synthesis technology as well as the selection of the entire TTS system in most researches depends on the target application and its requirements. Apart from research uses, synthetic speech is nowadays commonly used for example in toys, train announcements, telephone services or call centers, car navigation systems, mobile phones and computers.

At present, there are numerous existing TTS systems for other languages such as in Filipino (Guevara et al., 2001) (Corpuz et al., 2002), Mandarin (Chiang et al., 2002), Korean (Yoon, 2005), Amharic (Anberbir and Takara, 2009), Arabic (Khalifa, 2011), Portuguese (Oliviera et al., 1992), and others. For example, in Ethiopia, an Amharic Text-to-Speech system has been developed (Anberbir and Takara, 2009), which is parametric and rule-based that employs a cepstral method. The system uses a source filter model for speech production and a Log Magnitude Approximation (LMA) filter as the vocal tract filter. This Amharic TTS application is used by millions of people in Ethiopia since it is considered as the official language of the place. In Ethiopia, there are available TTS applications but only for Major language like English and Japanese. However, thousands of the world's 'minor' language lack technologies for TTS application and researchers in the area are quite very few. Therefore there is a strong demand to develop the TTS application for the African minor languages for the benefit of the African people especially in Ethiopia.

In the Philippines, the development of Filipino TTS system using Concatenative Synthesis is another existing TTS (Guevara et.al, 2001). Concatenative speech synthesis concatenates different length of pre-recorded speech samples obtained from natural speech. It requires less computational complexity at the expense of larger memory space. However, storage cost nowadays is quite inexpensive and so many are employing the concatenative speech synthesis in the field of Text-To-Speech. In the said study, a collaboration with the Department of Linguistics of the University of the Philippines – Diliman Campus was established to build a speech corpus. The system used phonemes as the basic acoustical units however the authors are still looking for the possibility of improving the synthesized speech through the use of syllables instead of phonemes. The TTS application is available for those researchers who are studying Natural Language Processing (NLP) and who want to enhance the said system.

Cebuano/*Bisaya* (Rubrico, 1998) is an Austronesian language spoken in the Philippines by about 20 million people. It is most widely spoken member of the Visayan language; a language closely related to Malaysian, Indonesian and the Filipino people. It is the second most spoken language in the Philippines. Cebuano is spoken in the provinces of Cebu, Bohol, Negros Oriental, western parts of Leyte, some parts of Samar, Negros Occidental, Biliran Island southern region of Masbate Island and Mindanao. Some dialects of Cebuano have different names to the language. Ethnic groups from Bohol may refer to Cebuano as "Bol-anon" while Cebuano speakers in Leyte identify their dialect as "kana".

Speakers in Mindanao refer to the language simply as "*Bisaya*". *Bisaya* has 20 phonemes. There are fifteen consonants: b, k, d, g, h, l, m, n, ng, p, r, s, t, w, and y. There are five vowels: a, e, i, o, u. *Bisaya* word has morphophonemic processes (Rubrico, 1998) which consist of the following: ***assimilation*** which occurs during affixation when a phoneme takes the point articulation of its neighbor; ***deletion*** which certain vowel or consonant are deleted after suffixation; ***alteration*** which alternates some consonant after suffixation and ***metathesis*** which is the process of reordering the phonemic sequence after suffixation which are the basis for the syllabification rules.

Many researches in Text-to-Speech have been done in recent years as mentioned above. However, these researches use different tools/synthesizers to come up with a synthesized speech depending on the language requirement. In this study, text analysis which includes tokenization of the input text, then separating the token into its syllables by following the set of rules in syllabification, recording the phoneme, diphone, triphone/special syllables as the speech database and the concatenation of pre-recorded speech were being studied to come up with a *Bisaya* Text-to-speech system. This system is relevant for *Bisaya*-speaking people, visually impaired person,as well as non-*Bisaya* speakers who want to know the *Bisaya* dialect. Moreover, this is a good start for *Bisaya* Natural Language Processing (NLP) researches particularly in Text-to-Speech system considering that *Bisaya* TTS is not yet available.

Methodology

General Architecture of Bisaya Text-to-Speech

Figure 1 presents the overall framework of *Bisaya* Text-to-Speech system. It illustrates the processes and procedures undertaken in the study. It consists of Input text – where the identified *Bisaya* words serve as an input. Based on the input string, the system identifies the token and each token is analyzed based on the rules identified for syllabification. Phoneme selection as well as the syllabification rule is implemented. Finally, the system checks from the database the corresponding speech for each syllable. The system then concatenates the pre-recorded speeches and gives the output as sound or speech.

Designing Rule-Based Algorithm

The general *Bisaya* dialect has been identified to be used as an input to the system. Interviews to some native *Bisaya* speakers were conducted to gather related information such as the commonly spoken *Bisaya* words. However, expressions made by a native *Bisaya* speaker in a particular place are not included in the study. Table 1 shows some sample *Bisaya* words that are commonly spoken.

Once the input text is provided, the system normalizes or processes the text which performs word or sentence tokenization. If the input text is a word, it is considered as one (1) token. The algorithm of the syllabification rules checks the token to syllabicate the words. If the input text is a sentence, the system identifies the tokens and each token evaluated based on the eleven (11) rules for syllabification.

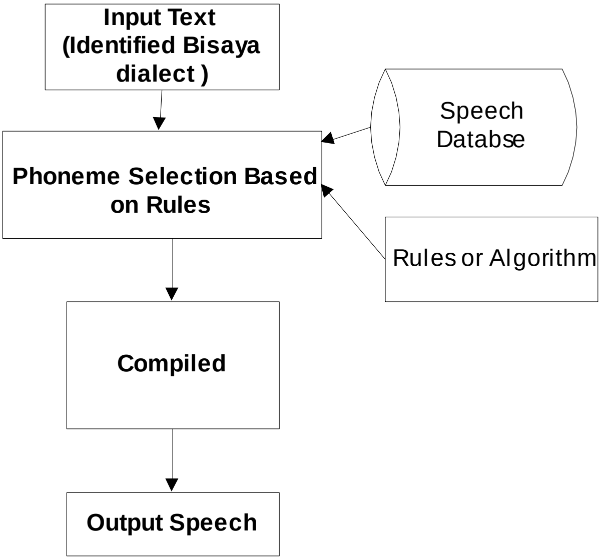

Input Text (Identified Bisaya dialect)
Speech Databse
Phoneme Selection Based on Rules
Rules or Algorithm
Compiled
Output Speech

Figure 1. General Architecture of *Bisaya* Text-to-Speech

Table 1. Sample *Bisaya* Words

Kaon	*Laba*	*Abot*
Ang	*Hagdanan*	*Pamatia*
Tinubdan	*Ingna*	*Plato*
Plangka	*Plantsa*	*Platito*
Nako	*Ko*	*Paniudto*
Kinahanglanon	*Binuhatan*	*Magtoon*
Paabota	*Hingtungdan*	*Kaugalingon*
Ginsakpan	*Ginsakupan*	*Nangabuang*
Kinabuang	*Gikabuangan*	*Pangasaba*
Badlungon	*Sabutanan*	*Istoryahon*
Paingon	*Hitsuraan*	*Buotanon*
Klaro	*Klarohon*	*Klarohonon*

Non- standard words like numbers, abbreviations and symbols are considered as invalid input. Some rules in *Bisaya* dialect are identified. With this, an algorithm is created that satisfies the needs which serves as the basis in concatenating pre-recorded speech. This step consists of tokenization and phoneme selection**.** First, the input *Bisaya* text is provided, after which the lexical analyzer identifies the different tokens in a given string. In order to identify each token, the spaces are considered as the identifier for a token. Non-standard words such as numbers, abbreviations and symbols are considered as invalid. Identified tokens are interpreted and selected based on the rules identified.

Bisaya phonemes are p, t, k, b, d, g, m, n, ng, s, h, w, l, r, and y (Rubico, 1998). There are five vowels: i, e, a, o, u. There are morphophonemic processes of a word; namely:

a.) Assimilation; which occurs during affixation when a phoneme takes the point of articulation of its neighbor like maN + kahoy > mangkahoy > mangangahoy > mangahoy.

b.) Deletion; in which the final vowel of the base may be deleted after suffixation like: Base = agad; Affix = + -on; For Deletion = agadon → "agdon".

c.) Alternation; in which it alters a certain phoneme like: [l] alternates with [w] when it is between /a/ and /u/ "balud > bawud", "bulad > buwad"; and,

d.) Metathesis, which is the process of reordering the phonemic sequence after suffixation, is often accompanied by the deletion of the final vowel of the base. For example; Base = tanum; Affixation = + -an; Deletion = tanuman; Metathesis = tamnan.

Based on the morphophonemic processes, rules were identified by the syllabification. These rules were then used for selecting the correct phonemes that were stored in a speech corpus. Table 2 presents the eleven (11) syllabification rules identified in *Bisaya* dialect and with corresponding example words. The underlined letters correspond to the syllabification rule.

Table 2. Syllabification Rules, Situations and Sample Words

Rule Number	Syllabification Rules and Situation	Sample Word(s)
1	**Vowel** When a vowel followed by a diphone or triphone, must be divided separately.	*(a, e, i, o, u) abot, ako, ibog*
2	**Vowel + ng** A vowel followed by an "ng" form one (1) syllable.	*Ang*
3	**Vowel + consonant** A vowel followed by a consonant.	*Ugma*
4	**Consonant + vowel** A consonant followed by a vowel.	*Laba*
5	**Consonant + vowel + consonant** A vowel between consonants.	*Bantog*
6	**Consonant + vowel + vowel** Two (2) consecutive vowels must be divided.	*Kaayo*
7	**Consonant + vowel + ng** A consonant followed by a vowel and "ng"	*Bingka*
8	**Consonant + consonant + vowel** A consonant cluster followed by a vowel	*klaro, grabe*
9	**Ng + vowel** An "ng" always followed by a vowel	*Ngalan*
10	**Ng + vowel + consonant** An "ng" plus a vowel followed by a consonant.	*Nganhi*
11	**Consonant + consonant + vowel + ng** A consonant cluster followed by a vowel and "ng" creates one (1) syllable	*Plangka*

Table 3 shows sample *Bisaya* words with syllabification and syllabification rules applied in each word. It must be pointed out that in *Bisaya* words, one (1) or more rules and syllabifications are being identified and applied as shown in the examples.

Table 3. Sample *Bisaya* Words, Syllabification and Syllabification Rules Applied

Sample *Bisaya* Words	Syllabification	Syllabification Rules Applied
Kaon	/ka/, /on/	Rule 4 and Rule 3
Ang	/ang/	Rule 2

Tinubdan	/ti/, /nub/, /dan/	Rule 4 and 2 Rule 5
Plangka	/plang/, /ka/	Rule 11 and Rule 4
Nako	/na/, /ko/	All Rule 4
Kinahanglanon	/ki/, /na/, /hang/, /la/, /non/	Rules4, 4, 7, 4 and Rule 5
Paabota	/pa/, /a/, /bo/, /ta/	Rule 6, 4 and Rule 4
Ginsakpan	/gin/, /sak/, /pan/	All Rule 5
Kinabuang	/ki/, /na/, /bu/, /ang/	Rule 4, 4, 4 and Rule 2
Badlungon	/bad/, /lu/, /ngon/	Rule 5, 4 and Rule 10
Paingon	/pa/, /i/, /ngon/	Rule 6 and Rule 10
Klaro	/kla/, /ro/	Rule 8 and Rule 4
Kaugalingon	/ka/, /u/, /ga/, /li/, /ngon/	Rules 6, 4, 4 and Rule 10
Pamatia	/pa/, /ma/, /ti/, /a/	Rules 4, 4 and Rule 6
Pangasaba	/pa/, /nga/, /sa/, /ba/	Rules 4, 9, 4 and 4
Buotanon	/bu/, /o/, /ta/, /non/	Rules 6, 4 and 5
Istoryahon	/is/, /tor/, /ya/, /hon/	Rules 3, 5, 4 and 5
Abot	/a/, /bot/	Rules 1 and 5
Paniudto	/pa/, /ni/, /ud/, /to/	Rules 4, 4, 3 and 4
Panihapon	/pa/, /ni/, /ha/, /pon/	Rules 4, 4, 4 and 5

Designing the Speech Database

In designing the speech database, the most common *Bisaya* words were identified. With a thorough evaluation of each word, phonemes, diphones, triphones and special syllables were also classified. These classifications (phonemes, diphones, triphones and special syllables) were then recorded in a semi-soundproof room using a microphone. The recording process was done by a professional speaker. The text corpus was read in a monotone voice having a normal pitch. The speeches were created and filed in a digital format (.wav) in any length. These pre-recorded speeches were utilized and employed in rule-based algorithm process and the basis for concatenation process to come up with an output voice or audio. The speech corpus were recorded through the use of a voice editing application called "Audacity" (see Figure 2).

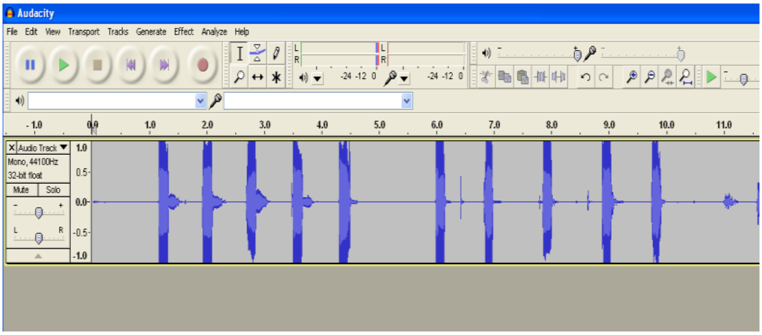


Figure 2. Recording Process Using Audacity Software

Hardware and Software Requirements

The hardware utilized in developing the *Bisaya* TTS system are the following: computer system which consists of the processor (dual-core) to process all the applications and programs of the *Bisaya* TTS system; memory (250GB HDD) for the storage of the pre-recorded speech and the keyboard or any input devices for input text; and the speaker for the speech output. The software used were the Java Programming Language (Java Package) which was utilized in coding the algorithm, user interface and concatenation processes; and, the Audacity software which is an audio editing software that was utilized in recording the speech and eliminating some noise.

Evaluation of the Bisaya TTS System

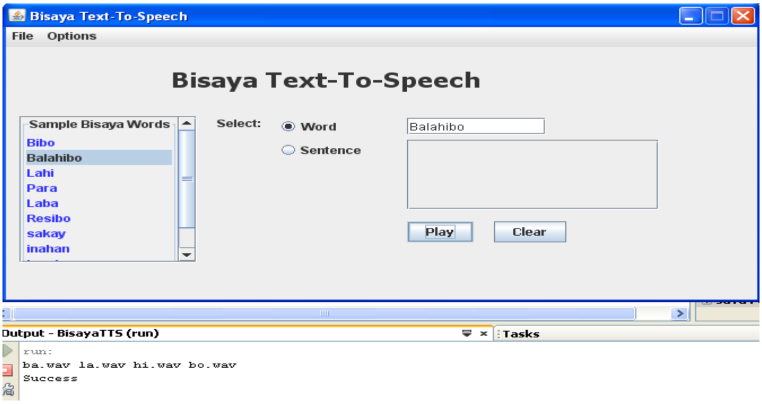
There were two (2) main criteria that are employed in evaluating the Bisaya TTS; namely, *Intelligibility and Naturalness.* Intelligibility is concerned with the ability to identify what was spoken or synthesized. It is less concerned with the naturalness of the signal, although naturalness inevitably related to and influences intelligibility. Naturalness is concerned on how close to human speech the output of the TTS system sounds. The well-known scoring method of these two (2) criteria is the *Mean Opinion Score* (MOS). The intelligibility and naturalness of the Bisaya TTS system were evaluated by comparing it with an existing English TTS system. In this study, the TextAloud, which is an existing system designed for English language, was used. TextAloud is a Text to Speech software for the Windows PC that converts text from MS Word Documents, Emails, Web Pages and PDF Files into natural-sounding speech. One can listen on PC or create audio files for use on iPods, iPhones, and other portable audio devices.

Listeners or respondents who are native *Bisaya* speakers of various level of education (students, professionals, non-professionals, no degree or no educational level) were gathered and were given instructions about the two TTS systems. Two sets were performed in two different TTS systems: (a) ten (10) Bisaya word samples, and, (b) t ten (10) Bisaya phrases/sentences. The respondents were asked to rate the speech quality of different systems, that is, the intelligibility and naturalness of the synthesis on a scale of 1-5, where 1 is Bad, 2 is Poor, 3 is Fair, 4 is Good and 5 is Excellent. All results were summed, and mean scores range from 1 to 5 were then derived, to represent the overall intelligibility and naturalness rating of the system.

Results and Discussion

User Interface Screen Shot

Figure 3 shows the user interface of *Bisaya* TTS system. It has 2 menus, namely: the File and Options. On the left part of the window are sample *Bisaya* words that is for easy access however, one can type the word on the “word” textbox. Sentence entry will be typed on the “sentence” textbox and the “play” button will then be clicked.



Select: ● Word ○ Sentence

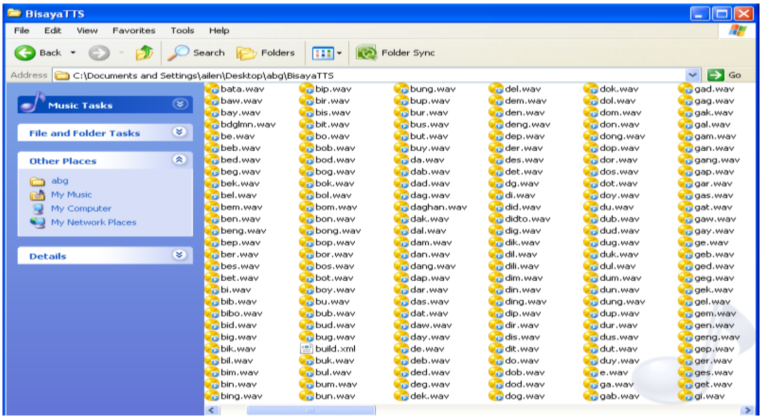
Balahibo

```
run:
ba.wav la.wav hi.wav bo.wav
Success
```

Figure 3. *Bisaya* TTS System User Interface Screen Shot

Database of Pre-recorded Speech

Figure 4 illustrates the list of recorded speech in one folder. It is consists of phonemes, diphones, triphones and special syllables. Around 1,100 speeches were recorded. The recorded speeches were in (.wav) file extension. The filename for each speech must match the syllables identified considering the 11 rules based on the inputted text. Each of the speech has different length and filename and this will be the basis of the system for appending or concatenation. However, the frequency is normal, that is, there is no low or high pitch since the recording process followed the monotonic voice in a continuous file and each syllable was then cut and saved with a separate file with corresponding filename, as shown in Figure 4.



Address: C:\Documents and Settings\ailen\Desktop\abg\BisayaTTS

```
bata.wav     bip.wav      bung.wav     del.wav      dok.wav      gad.wav
baw.wav      bir.wav      bup.wav      dem.wav      dol.wav      gag.wav
bay.wav      bis.wav      bur.wav      den.wav      dom.wav      gak.wav
bdglmn.wav   bit.wav      bus.wav      deng.wav     don.wav      gal.wav
be.wav       bo.wav       but.wav      dep.wav      dong.wav     gam.wav
beb.wav      bob.wav      buy.wav      der.wav      dop.wav      gan.wav
bed.wav      bod.wav      da.wav       des.wav      dor.wav      gang.wav
beg.wav      bog.wav      dab.wav      det.wav      dos.wav      gap.wav
bek.wav      bok.wav      dad.wav      dg.wav       dot.wav      gar.wav
bel.wav      bol.wav      dag.wav      di.wav       doy.wav      gas.wav
bem.wav      bom.wav      daghan.wav   did.wav      du.wav       gat.wav
ben.wav      bon.wav      dak.wav      didto.wav    dub.wav      gaw.wav
beng.wav     bong.wav     dal.wav      dig.wav      dud.wav      gay.wav
bep.wav      bop.wav      dam.wav      dik.wav      dug.wav      ge.wav
ber.wav      bor.wav      dan.wav      dil.wav      duk.wav      geb.wav
bes.wav      bos.wav      dang.wav     dili.wav     dul.wav      ged.wav
bet.wav      bot.wav      dap.wav      dim.wav      dum.wav      geg.wav
bi.wav       boy.wav      dar.wav      din.wav      dun.wav      gek.wav
bib.wav      bu.wav       das.wav      ding.wav     dung.wav     gel.wav
bibo.wav     bub.wav      dat.wav      dip.wav      dup.wav      gem.wav
bid.wav      bud.wav      daw.wav      dir.wav      dur.wav      gen.wav
big.wav      bug.wav      day.wav      dis.wav      dus.wav      geng.wav
bik.wav      build.xml    de.wav       dit.wav      dut.wav      gep.wav
bil.wav      buk.wav      deb.wav      do.wav       duy.wav      ger.wav
bim.wav      bul.wav      ded.wav      dob.wav      e.wav        ges.wav
bin.wav      bum.wav      deg.wav      dod.wav      ga.wav       get.wav
bing.wav     bun.wav      dek.wav      dog.wav      gab.wav      gi.wav
```

Figure 4. List of Recorded Speech (database)

Evaluation of the Bisaya TTS System

To be able to measure or evaluate the intelligibility and naturalness (Scansoft, 2004) of the *Bisaya* TTS system, it should be compared and tested with an existing TTS system . In this study, the *Bisaya* TTS is compared with the *TextAloud* which is an existing system designed for English language. *TextAloud* is Text to Speech software for the Windows PC that converts the text from MS Word Documents, Emails, Web Pages and PDF Files into natural-sounding speech. Once can listen in the PC or create audio files for use on iPods, iPhones, and other portable audio devices.

Listeners or respondents who are native *Bisaya* speakers of various level of education who are (students, professionals, non-professionals, no degree or no educational level) were gathered and were given instructions about the two TTS systems. Two sets were performed in two different TTS systems: (a) ten (10) Bisaya word samples, and, (b) the ten (10) Bisaya phrases/sentences. The respondents were asked to rate the speech quality of different systems, that is, the intelligibility and naturalness of the synthesis on a scale of 1-5, where 1 is Bad, 2 is Poor, 3 is Fair, 4 is Good and 5 is Excellent. The mean scores range from 1 to 5 were then derived, which represent the overall intelligibility and naturalness rating of the system.

Table 4 shows the results of the evaluation of Mean Opinion Score rating for "word". The *Bisaya* TTS system has the higher rating with equivalent compared to *TextAloud* TTS system. Two words garnered a rating of 5 (Excellent) and rest have a rating of 4 (Good) for the *Bisaya* TTS system. On the other hand, majority of the words under the *TextAloud* TTS system got a rating of 1, which is equivalent to "Bad". Hence, the *Bisaya* TTS system is more intelligible and natural as compared with that of using the *TextAloud* TTS system.

Table 4. Evaluation Result of the Two Different TTS Systems for Word

Mean Opinion Score (MOS) Rating for Word				
***Bisaya* Words**	***Bisaya* TTS**		***TextAloud* TTS**	
	Rating	**Equivalent**	**Rating**	**Equivalent**
Hagdan	4.2	Good	2.2	Poor
Balahibo	4.3	Good	1.4	Bad
Para	4.0	Good	1.1	Bad
Resibo	4.1	Good	2.3	Bad
Bibo	4.1	Good	1.0	Bad
Inahan	3.9	Good	1.0	Bad
Lahi	4.8	**Excellent**	1.0	Bad
Laba	4.8	**Excellent**	1.0	Bad
Ngalan	4.0	Good	1.0	Bad
Paabot	3.5	Fair	1.0	Bad

In Table 5, the evaluation results of the Mean Opinion Score rating for "phrase /sentence are shown. *Bisaya* and *TextAloud* TTS systems were evaluated using a phrase or sentence as an input text to the system. Similar with the results for "word", the *Bisaya* TTS system gains higher mean opinion score rating. Hence, *Bisaya* TTS system is more intelligible and natural where people can understand the output sound or speech.

Table 5. Evaluation Result of Two Different TTS Systems for Phrase/Sentence

Mean Opinion Score (MOS) Rating for Phrase/Sentence				
Bisaya Phrase/Sentence	***Bisaya* TTS**		***TextAloud* TTS**	
	Rating	**Equivalent**	**Rating**	**Equivalent**
Nahulog ang resibo sa hagdan.	3.5	Fair	1.1	Bad
Labong kaayo iya balahibo sa tiil.	3.5	Fair	2.2	Poor
Ginoo pamatia kami.	3.5	Fair	1.0	Bad
Bibo kaayo ang pasko.	3.8	Good	1.4	Bad
Ang ako inahan kay buotan.	3.8	Good	1.0	Bad
Nagsakay siya ug baka.	3.6	Good	2.1	Poor
Gipalong niya ang plangka sa kuryente.	3.6	Good	2.8	Fair
Nagpaabot sila sa balay.	3.5	Fair	2.2	Poor
Bangon Kagay anon.	3.5	Fair	1.1	Bad
Daghan ang nangamatay sa baha.	3.6	Good	1.1	Bad

Conclusions and Recommendation

This study developed a *Bisaya* TTS system utilizing rule-based algorithm and concatenative speech synthesis. It utilized the set of rules in syllabification as the created algorithm and employed pre-recorded speech database which are the basis for concatenation. Evaluation of the *Bisaya* TTS system shows that it is more intelligible for the native *Bisaya* speakers compared to existing *TextAloud* TTS system.

For future work, it is worth exploring to find more sophisticated model for the prosodic part to come up with very natural utterances and intelligibility. The normalization of the speeches, the different stress and the intonation of each word and type of sentences should be considered to avoid confusions.

REFERENCES

Anberbir, Tadesse and Takara, Tomio, "Development of an Amharic Text-to-Speech Using Cepstral Method", *Proceedings of the First Workshop on Language Technologies for African Languages* (2009): 46-52.

Chiang, Fu; Tseng, Chiu-Yu and Lee, Lin-Shan, "A Set of Corpus-Based Text-to-Speech Synthesis Technologies for Mandarin Chinese", *IEEE Transaction on Speech and Audio Processing*.10; 7(2002):1- 9.

Corpus, M.; Liampo, J.; Co, And Guevara, R.C. L, (2004) "Development of a Filipino TTS System Using Concatenative Speech Synthesis", *Proceedings of 2nd National ECE* Conference (2001).

Guevara, R.C.L.; Co, M., Tan,E; Garcia, I.D.; Espina, E.; Ensomo, R. and Sagum, R., "Development of a Filipino Speech Corpus", *Proceedings of 3rd National ECE Conference* (2002).

Khalifa, Othman,"A Rule-Based Arabic Text-to-Speech System Based on Hybrid Synthesis Technique", *Journal of Basic and Applied Sciences*, 5; 6 (2011): 342-354

Oliveria, L.C, Viana M.C. and Trancoso, I.M., "A Rule-Based Text-to-Speech System for Portuguese", *Acoustics, Speech, and Signal Processing,* 2 (1992):73-76.

Rubrico, Jessie Grace, *Cebuano Grammar Notes,*Language Links ,1998.

Schroeder, M. R. (1993). "A Brief History of Synthetic Speech", Speech Communication, 13 (1993): 231-237.

Text-to-Speech (TTS), Accessed July 2011. http://searchmobilecomputing.techtarget.com/ definition/text-to-speech

Yoon, Kyuchul, Building a Prosodically Sensitive Diphone Database for a Korean Text-to-Speech Synthesis System, (PhD diss., Ohio State University, 2005).