Designing the balinese script-to-speech synthesis system using noto serif balinese font

M H Setiawan¹, M W Aditya², L H Loekito³ and G Indrawan⁴

¹,²,³Computer Science Department, Universitas Pendidikan Ganesha, 81116, Indonesia
⁴Electrical Engineering Department, Universitas Pendidikan Ganesha, 81116, Indonesia

Email: heri.setiawan@undiksha.ac.id

Abstract. The existence of regional languages shows an identity of the existence of a tribe. Indonesia has many regional languages, one of which is Balinese with Balinese script as input for the script which comes from Sanskrit. Reading Balinese script requires knowledge of Balinese script and Balinese culture in order to be able to correctly pronounce the written Balinese script. Since learning Balinese script took a long learning process, it certainly causes a slowdown in knowledge about how to read the Balinese script. Therefore, it is necessary to have a system that can translate Balinese script into a speech so that it can be used in learning the Balinese script without any mistakes in pronouncing vowels. This research designed a system synthesis of Balinese script into a speech with the system synthesis method of Noto Serif Balinese Font which consists of three processes namely text pre-processing, prosody generation, and concatenation. Concatenation or combining speech will use the Pitch Synchronous Overlap Add Method (PSOLA). The result to be achieved in this research is to be able to decompose the Unicode from the Noto Serif Balinese font into a sentence which then this sentence will become a message. In the future, this research can use input from an image to directly enter into speech. Making it easier for users to translate the Balinese script.

1. Introduction
Balinese language is one of the regional languages that has its own script which is called the Balinese script. The development of Balinese Script has slowed down due to the reading of Balinese Script that requires the knowledge of Balinese Script and Balinese Culture which is not brief. This is also influenced by the younger generation who are starting to abandon the use of Balinese script [1] which can lead to the extinction of Balinese script and the decline of Balinese culture. The effort to preserve Balinese culture, including the use of Balinese language and Balinese script, is stated in the Governor of Bali Regulation Number 80 of 2018 concerning the Protection and Use of Balinese Language, Script and Literature [2] [3] which is addressed to the government and non-government institutions. When reading the Balinese script, the pronunciation or pronunciation of different vowels will have different meanings. To help preserve and help accelerate the learning of reading Balinese script, this research will try to develop a system that can translate Balinese script into speech, so that it is hoped that there will be no mispronunciation that results in misinterpretation in reading Balinese script. Previous research on transliteration [4] by Indrawan, Paramarta, and Agustini in 2019 has developed web and mobile-based applications to transliterate from Latin text into the Balinese script using Noto Sans Balinese letters [5] and Loekito et al. [6] have provided development and success increase accuracy by up to 94% with the
addition of special words and corrections. More developed dedicated Balinese-Unicode font, was found on Noto Serif Balinese font rather than Noto Sans Balinese font [7] [8] [9] [10]. This study aims to develop this application that previously could translate Latin text into the Balinese script as well as aiming to also have function in reading the results of the transliteration into a speech.

2. Literature review
2.1. Balinese Script
Balinese script is a manuscript from the ancient Indian Brahma manuscript. The manuscript has similarities with modern manuscripts from Southeast Asia and South Asia which are also derived from the Brahma script [11]. Reference [12] gives classification to Balinese script into several parts, namely: 1) basic syllables, 2) sound characters consisting of vowels, independent vowels, illegal syllable combinations, and semi-vowels, 3) Swalalita script, 4) Pangangge Tengenan, and 5) Numbers, and other signs.

Pronunciations in Balinese script is particular concern due to the different pronunciation of the vowels a and å, i with ĩ, e with ē, u with ū, and o with ő where different pronunciations produces different meanings.

2.2. Noto Serif Balinese Font
Noto Serif Balinese font is a font family that is integrated with Google Noto Fonts, which includes various languages, both Latin and Manuscript Languages [13] [14]. Noto Serif Balinese font is designed based on the Unicode standard. The Unicode standard is a standard for universal character encoding rules for text and written characters [15]. The image below shows the character map that is in Noto Serif Balinese font.

![Figure 1. Unicode noto serif balinese font](image)

2.3. Phoneme
The phoneme is a linguistic term and is the smallest unit in a language that can still show different meanings [17]. Balinese script has a phoneme that is caused by Aksara Swara and Aksara Swalalita. According to reference [18], phonemes are a collection of types of language that have a speech sound, where the sound of the speech serves to provide a difference in the meaning of the word (distinctive). A study in phonology explains that sounds or fonts are written in brackets [ ], phonemes are written in two slashes //, and letters between brackets ( ).

2.4. Diphone
Diphone is a combination of two phonemes in the form of the sound [17]. The basic syllable of the Balinese script is included in one of the diphones when the basic syllable is voiced.

2.5. Systems Text to Speech Synthesis
Text to speech is a system that can convert text into speech. Two principles occur during the processing of text to speech, namely: the text to phoneme converter part, and the phoneme to speech converter part. The first part serves to convert the input text into a series of sound codes which are usually represented by phoneme codes, duration, and pitch.
The second part serves as a converter from phoneme to speech. In this process, the input that will be received is in the form of phoneme codes as well as the pitch and duration produced by the previous section.

The change from text to phoneme is greatly influenced by the rules that applied in a language. The results in the implementation of the text to phoneme converter part is being specific to a language.

![Figure 2. The order of conversion from text to speech.](image)

In general, synthesis is determined by 3 characteristics, namely [18]:
1. The level of understanding (intelligibility), namely the speech from the conversion of text into speech can be understood by listeners.
2. Fluidity, that is, the pronunciation of the converted pronunciation is fluent.
3. The naturalness of the pronunciation (naturalness), that is, the pronunciation of the conversion result sounds according to human pronunciation in general.

2.6. Pitch Synchronous Overlap Add Method
The basic idea of PSOLA is to decompose the speech signal into individual pitch periods so that we can move those pitch periods to be replaced by the effective length of the pitch periods. Further, the fundamental frequency of a signal will be expressed as a periodic structure concerning the time. PSOLA is also an algorithm variation of SOLA (Synchronous Overlap-Add) [19].

The PSOLA method is a method of cutting the signal into segments according to the length of the periodic structure, then the result of the slice of the signal can be placed as desired and can be added with other signal segments. This adding process will look superimposed. Since the short-term correlation in the signal does not change, that is, the signal in the segment does not change, the mix of the signal spectrum does not change. The following is an overview of the PSOLA method [20]:

![Figure 3. PSOLA method process [20].](image)

3. Methods & System Design
The text to speech system method consists of 3 parts of the process, namely text pre-processing, process generation, and concatenation, as well as speech recording, design systems, user interface. Below is a block diagram of a text to speech system design:
3.1. Text pre-processing

This process is a process of converting the Balinese script into a diphone. When the input is in the form of basic characters, numbers, voice characters, self-help characters, this section will convert them to diphones that are already available in the diphone database. The block diagram in this process is as follows:

![Text Pre-processing Block Diagram](image)

- **Text Aksara**
  In this section, the Balinese script that has been entered into the system is checked so that later the input will be checked whether it is Balinese script with the Noto Serif Balinese font type. Examples of character text input are as follows: (Kala: time).

- **Get Unicode**
  This process is the conversion of the manuscript fonts into Unicode fonts. Further, Unicode sequence in 1 word or 1 sentence will be received. The example could be seen from the Kala sentence above where Unicode 1B13 (Ka) 1B2E (La) was received.

- **Converter Number**
  The system will convert Unicode which is a number into a phoneme representation if the input of the Script text is a number.

- **Segmented Word**
  If the input in the system is a word or sentence, the system converts the word or sentence into a diphone representation. The example could be seen from the word Kala above where the result shows Unicode 1B13 1B2E with the help of dictionary phonemes further resulting the phoneme /ka/la/.

- **Phoneme Dictionary**
  It is a database consisting of a collection of Balinese script phonemes. The data structure of the phoneme dictionary is a database table with 3 columns namely Unicode column, phoneme column, and character type column. Unicode column contains all Unicode from the Noto Serif Aksara Balinese letters, the phoneme column contains the phoneme representation of Unicode Noto Serif Aksara Bali, while the script type column contains the script type according to reference [12].
• **Multi-Level Data Structure (MLDS)**
  Consists of all the data needed for the next subsystem namely the prosody process. MLDS consists of drawing diphones which are converted into inputs. MLDS is required for the process of prosody formation. MLDS consists of words, diphone representation, and prosodic parameters for each diphone. Using the previous example with Kala, you can get MLDS data [kala] (word), -k/ka/la/a- (diphone representation), and the prosodic parameters are the addition of the final pitch diphone /-k/, the addition of the initial and final pitch for diphone / ka / la /, and 1 pitch prefix increments on diphone / a- /.

3.2. **Prosody Generate**
This process of prosody is a process to get a more natural speech because prosody is a process of giving intonation of a speech. In quantitation, prosody is the change in the pitch value (fundamental frequency) during the pronunciation of the sentence or pitch as a function of time. The use of intonation in speech is very specific for each language, so the model needed to generate prosody data is very specific for a language. The following is a block diagram of this prosody process:

![Prosody block diagram](image)

- **MLDS**
  Consists of all the data that has been stored in the previous process namely the word, diphone description, and parameter prosody for each diphone. The parameter in question is the combination of the word level and the sentence prosody level.
- **Diphone Retrieval**
  In this process, there are three steps that must be carried out namely the diphone recording database, storing the diphone waveform and prosodic parameters in variables. Referring to the previous example, the diphone -k / ka / la / a- will be taken from the diphone database and further the pitch requirements were added according to the data provided by the MLDS process.
- **Acoustic Manipulation**
  This process includes recognition of .wav wave files (load, play, write), a vast array of signal processing equipment obtained according to data from the diphone retrieval process.

3.3. **Concatenation**
The concatenation process is the merging of previously recorded sound segments. The diphone that has been processed in the prosody generate will be reprocessed in this process, that is, each additional pitch in the previous process will be windowed. This function is to reduce the discontinuity effect at the edges of the frame which is generated during the framing process. The window function used is the hamming window, namely:

\[
W(i) = 0.54 - 0.46 \left(1 - \cos \left(\frac{2\pi i}{n}\right)\right), 0 \leq i \leq n - 1
\]  

(1)
3.4. Speech Recording

In this system, a phoneme database is needed, namely the speech unit formed from Balinese script phonemes. The difference in phoneme form depends on the location in the word whether it is at the beginning of the syllable, middle, or at the end [18]. The recordings will be saved into the diphone database with the .wav format.

3.5. Design Systems

At this stage, the software is designed that will receive input in the form of Balinese script which will be automatically converted by the system into a phoneme that involves prosody using Balinese intonation. Furthermore, the resulting phoneme will be inputted for applications that implement PSOLA. The general description of how the application will be designed is made in the form of object-oriented modelling such as use cases, sequences, activities, and class diagrams.

3.6. User Interface

The Graphic user interface (GUI) is designed to make it easier for users to interact with the system and test the methods used. The GUI design phase will refer to reference [5] which will be developed with the addition of methods in this study.

4. Result and Discussion

Based on the results of the study, all of the steps including making software in the block method and system design, testing which was carried out by entering words, sentences, or numbers, as well as transliterating into Balinese script. Further it then enters Balinese script which will be input for the system and processed into pre-processing text blocks. The words or sentences that were inputted will be converted into phoneme and diphone representations. If number is inputted, the system will convert it to a string. From this string form, it will be converted into a diphone representation like the example below.

From the series of processes above, if the input number is "123", the system will convert it into a string "seratus dua puluh tiga". From this string formation, conversion is carried out into the form of a diphone representation to become ".s / er / ra / t / tus / s- /, -d / du / a- /, -p / pu / ul / u / uh / h- /, -t/ti/ /a- ".

If the input is in the form of a word or sentence, the system will immediately convert the input into a phoneme representation form. An example of the word "Kadep" which is composed of diphone-diphone becomes "k / ka / d / de / ep / p-".

![Figure 7. Text pre-processing with word input.](image-url)
The next process, the diphone-diphone representation that has been obtained will be modelled in the generation of prosody. Each of the diphones will be added 1 pitch period of data at the front and back. It aims to apply the PSOLA algorithm (Pitch Synchronous Overlap Add Method) for the concatenation process.

The important thing in this process is determining the length of the data from the initial and final pitches because this determines how much data is overlapped and added to the original diphone. This is also the basis that is used to connect a diphone using PSOLA.

Some of the basic principles of the PSOLA method are:
- If the speech signal is most advanced, only one pitch period in the end will be processed.
- If the speech signal is in the middle, then the start and end of a pitch period will be processed.
- If the speech signal is the last one, then only the initial pitch period will be processed.

4.1. System Testing

System testing will be carried out using MOS (Mean Opinion Score) testing where the test is carried out on some listeners who will immediately listen to the synthesized speech signal and provide an assessment with the weights/criteria described in Table 1. The test will be carried out on 50 listeners, each listener will be given 10 different kinds of speech words. The 10 kinds of words used for testing are as follows: Kala, Kādep, Akṣara, Kērēng, Pakraman, Gaṇitri, Bangkung, Karṇa, Śiwa, and Upacāra.

| No | Criteria | Quality (q) | Weight (w) | Description |
|----|----------|-------------|------------|-------------|
| 1  | Intelligibility | Fair | 3 | Speech can be understood, the sequence of speech is quite clear |
|    |          | Good       | 4 | Speech can be understood, the sequence of speech is clear |
|    |          | Excellent  | 5 | Speech can be understood, the sequence of speech is very clear |
|    |          | Bad        | 1 | Speech is not smooth, transitions between syllables are very annoying |
|    |          | Poor       | 2 | Speech is not smooth, transitions between syllables are annoying |
| 2  | Fluidity | Fair       | 3 | Speech is quite fluent, transitions between syllables are a little annoying |
|    |          | Good       | 4 | Smooth speech, comfortable transitions between syllables |
|    |          | Excellent  | 5 | Smooth speech, transitions between syllables are very convenient |
|    |          | Bad        | 1 | Flat pronunciation with no marks |
|    |          | Poor       | 2 | The pronunciation is slightly toned, not according to human pronunciation in general |
| 3  | Naturalness | Fair     | 3 | The pronunciation is slightly toned, according to human pronunciation in general |
Pronunciation with good tones, according to human pronunciation in general

| No | Criteria | Quality (q) | Weight (w) | Description |
|----|----------|-------------|------------|-------------|
| 1  | Excellent| 5           |            | Identic pronunciation with common human pronunciation |

The results of filling out the assessment from the listeners will be processed by finding the average value using the following formula:

\[ MOS_{Intelligibility,Fluidity,Naturalness} = \frac{1}{50} \sum_{i=1}^{n} q_i w_i \]  

where \( i = \text{Bad}, \text{Poor}, \text{Fair}, \text{Good}, \text{Excellent} \), \( q = \text{quality} \) and \( w = \text{weight} \).

### 5. Conclusion

The addition of the text to speech feature that is applied in Balinese Noto Serif can be used to pronounce Balinese script that has been formed in transliteration, so that the pronunciation of the Balinese script does not experience pronunciation errors. Besides, this application can be used as a way to learn Balinese script pronunciation for students or those who want to learn the Balinese script. Based on the design and discussion described in this research, it can be concluded that the Balinese to Speech script system is a feature in the Noto Serif Balinese application which is expected to be useful in assisting the pronunciation of Balinese script to minimize mispronunciation. The availability of this feature in the previous application will add functions of the previous application, namely transliteration to transliteration and reciting the Balinese script.

Suggestions for further research are expected that the input from the system can be developed into an image recognition, where input can read Balinese script from an image so that it does not depend on Unicode from Noto Serif Balinese.

### Acknowledgments

This work was supported by Universitas Pendidikan Ganesha through the research grant No. 761/UN48.16/LT/2020.

### References

[1] Sudana A K O, Wibawa K S and Tirtha I M A D 2016 Learning Media Of Balinese Script Writing Based On Augmented Reality *Journal Of Theoretical and Applied Information Technology* 90 31-39

[2] Gubernur Bali 2018, *Peraturan Gubernur Bali No. 80 tentang Pelindungan dan Penggunaan Bahasa, Aksara, dan Sastra Bali serta Penyelenggaraan Bulan Bahasa Bali* [Bali Governor Regulation No. 80 on Protection and Usage of Balinese Language, Script, and Literature, also Organizin, Denpasar].

[3] Gubernur Bali 2019, *Surat Edaran Gubernur Bali No. 3172 Tahun 2019 tentang Penggunaan Busana Adat Bali dan Aksara Bali* [Bali Governor Circular Letter No. 3172 Year 2019 about The Usage of Balinese Traditional Clothing and Balinese Script], Denpasar.

[4] Jampel I N, Indrawan G and Widiana I W 2018 Accuracy Analysis of Latin-to-Balinese Script Transliteration Method *International Journal of Electrical and Computer Engineering (IJECE)* 8 1788-1797

[5] Indrawan G, Paramarta I K and Agustini K 2019 A New Method of Latin-to-Balinese Script Transliteration based on Noto Sans Balinese Font and Dictionary Data Structure *ICSIM 2019: Proceedings of the 2nd International Conference on Software Engineering and Information Management* pp 75-79
[6] Loekito L H, Indrawan G, Sariyasa and Paramarta I K 2019 Error Analysis of Latin-to-Balinese Script Transliteration Method Based on Noto Sans Balinese Font," in Proceedings of the 3rd International Conference on Innovative Research Across Disciplines (ICIRAD 2019), Singaraja

[7] Indrawan G, Ariawan K U, Agustini K and Paramarta I K 2020 Finite-State Machine for Post-Processing Method of Balinese Script to Latin Transliteration in The 6th International Conference on Science, Technology, and Interdisciplinary Research, Lampung

[8] Indrawan G, Setemen K, Sutaya W and Paramarta I K 2020 Handling of Line Breaking on Latin-to-Balinese Script Transliteration Web Application as Part of Balinese Language Ubiquitous Learning in The 6th International Conference on Science in Information Technology (ICSItech), Palu

[9] Indrawan G, Dantes G R, Aryanto K Y E and Paramarta I K 2020 Handling of Mathematical Expression on Latin-to-Balinese Script Transliteration Method on Mobile Computing in The 5th International Conference on Informatics and Computing (ICIC).

[10] Indrawan G, Gunadi I G A, Gitakarma M S and Paramarta I K 2021 forthcoming Latin to Balinese Script Transliteration: Lessons Learned from Computer-based Implementation in The 4th International Conference on Software Engineering and Information Management (ICSIM), Yokohama

[11] Habibi I and Rinaldi 2006 The Balinese Unicode Text Processing Indonesian Journal of Innovation in Soft Computing and Cybernetic Systems 11 1-17

[12] Sudewa I B A 2003 "babadbali.com/," 14 05 2003 [Online] Available: babadbali.com/aksarabali/alphabet-c.htm. [Accessed 23 09 2020].

[13] "Google Noto Fonts," 2014 [Online]. Available: https://www.google.com/get/noto/. [Accessed 30 09 2020].

[14] Google, "Google Noto Serif Balinese," 2020. [Online] Available: https://github.com/googlefonts/noto-fonts/blob/master/unhinted/ttf/NotoSerifBalinese/NotoSerifBalinese-Regular.ttf. [Accessed 24 10 2020].

[15] Narendra B B 2008 Pembuatan Smart Font Aksara Bali Dengan Graphite Description Language (Bandung: Institut Teknologi Bandung)

[16] Indrawan G, Puspita N N H, Paramarta I K and Sariyasa S 2018 LBtrans-Bot: A Latin-to-Balinese Script Transliteration Robotic System based on Noto Sans Balinese Font Indonesian Journal of Electrical Engineering and Computer Science 12

[17] Rachma H D, Rugniaga Z and Huda M 2011 Pembuatan Text-To_Speech Synthesis System Untuk Penutur Berbahasa Indonesia Electronic Engineering Polytechnic Institute of Surabaya (EPPIS) 13

[18] Melangi S 2018 Text To Speech Bahasa Indonesia Menggunakan Synthesizer Concatenation Berbasis Fonem Jurnal Teknik Elektro CosPhi, 2 31-36

[19] Bhaskoro S B, Ariani I and Alamsyah A A 2014 Transformasi Pitch Suara Manusia Menggunakan Metode PSOLA TELKOMNIKA 2 129

[20] Bäckström T 2020 "Pitch-Synchronous Overlap-Add," Aalto University Wiki, 08 2020. [Online]. Available: https://wiki.aalto.fi/pages/viewpage.action?pageId=155477136. [Accessed 8 10 2020].