

**SKRIPSI**

**Pembuatan *Corpus* Berbahasa Indonesia Menggunakan Metode *Diphone*  
*Concatenation* Untuk Sistem *Text-to-Speech***

**Disusun dan diajukan oleh**

**DHINDA FITRI WILUDJENG**

**D421 16 501**



**DEPARTEMEN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS HASANUDDIN**

**MAKASSAR**

**2022**

**LEMBAR PENGESAHAN SKRIPSI**  
**PEMBUATAN *CORPUS* BERBAHASA INDONESIA**  
**MENGGUNAKAN METODE *DIPHONE CONCATENATION* UNTUK**  
**SISTEM *TEXT-TO-SPEECH***


Disusun dan diajukan oleh  
**DHINDA FITRI WILUDJENG**  
**D421 16 501**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 1 Desember 2022 dan dinyatakan  
: telah memenuhi syarat kelulusan.

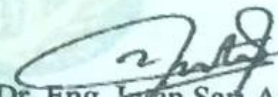
Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Dr. Indrabayu, S.T., M.T., M.Bus.Sys.  
Nip. 19750716 200212 1 004



Dr. Eng. Inan Sari Areni, S.T., M.T.  
Nip. 19750203 200012 2 002

Ketua Program Studi,



Dr. Indrabayu, S.T., M.T., M.Bus.Sys.  
Nip. 19750716 200212 1 004

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Dhinda Fitri Wiludjeng

NIM : D421 16 501

Departemen : Teknik Informatika

Jenjang : S1

Menyatakan dengan ini karya tulisan saya berjudul:

“Pembuatan *Corpus* Berbahasa Indonesia Menggunakan Metode *Diphone Concatenation* Untuk Sistem *Text-to-Speech*”

Adalah karya tulisan saya sendiri dan bukan merupakan pengambil alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 1 Desember 2022

Yang menyatakan,



Dhinda Fitri Wiludjeng

## KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Segala puji dan syukur kami panjatkan atas kehadiran Allah SWT Tuhan Yang Maha Esa pemilik semesta alam karena dengan limpahan rahmat penulis diberikan berkat, kekuatan, kesabaran dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul **“Pembuatan *Corpus* Berbahasa Indonesia Menggunakan Metode *Diphone Concatenation* Untuk Sistem *Text-to-Speech*”** ini dapat diselesaikan sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 pada Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini banyak mendapatkan sumbangan pikiran serta bimbingan baik moral maupun materil dari berbagai pihak. Sehingga pada kesempatan ini penulis menyampaikan ucapan terima kasih sedalam-dalamnya kepada:

1. Orang tua penulis yaitu Bapak Aris Susanto dan Ibu Sunarsih dan adik-adik yang tidak pernah lelah mendoakan dan memberi dukungan kepada penulis dalam mengerjakan penelitian ini.
2. Bapak Dr. Indrabayu, S.T., M.T., M.Bus.Sys., IPM selaku dosen Pembimbing I dan Ibu Dr.Eng. Intan Sari Areni, ST, MT. selaku pembimbing II, yang telah banyak memberi bimbingan, motivasi dan masukan yang bermanfaat untuk penulis.

3. Bapak Dr. Indrabayu, S.T., M.T., M.Bus.Sys., IPM Sebagai Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.
4. Staf Departemen Teknik Informatika Universitas Hasanuddin.
5. Teman-teman, adik-adik dan kakak-kakak dari AIMP Family yang telah memberikan begitu banyak bantuan selama penelitian, pengambilan data dan diskusi progress penyusunan Tugas Akhir.
6. Saudara seangkatan dari IGNITER16 yang senantiasa memberikan dukungan dan kebersamaan dalam menjalani perkuliahan.
7. Teman-teman seperjuangan yaitu AIROKU tercinta (Sitti, Sari, Ani, Marimar, Syukri) yang selalu ada membantu, memberikan dorongan dan selalu ingin sukses bersama.
8. Dan kepada Calonnya saya M. Rudini Kurniawan Amiruddin, S.ST., M.T yang senantiasa memberikan dukungan serta selalu sabar menghadapi penulis dalam keadaan apapun.

Penulis menyadari masih terdapat kekurangan dalam penyusunan Laporan skripsi, oleh karena itu penulis mengharapkan adanya saran dan kritik untuk pengembangan lebih lanjut terhadap penelitian selanjutnya. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan bagi penulis khususnya.

Makassar, 1 Desember 2022

Dhinda Fitri Wiludjeng

## ABSTRAK

Perkembangan *Natural Language Processing* (NLP) di Indonesia sangat terhambat karena kurangnya sumber daya seperti *Corpus*. Hal itu disebabkan karena kurangnya perhatian pemerintah terhadap bidang kebahasaan dan penelitian ini butuh waktu yang lumayan lama karena jumlah kata yang banyak dan beragam untuk membuat sebuah *Corpus*. Oleh karena itu, maka sistem ini dibuat agar dapat memberi kontribusi pada penelitian ke-Bahasa dan dapat membantu dalam pengembangan penelitian di bidang NLP di Indonesia dengan membuat sistem *Text-To-Speech*. Adapun metode yang digunakan untuk membangun sistem ini adalah metode *Diphone Concatenation*. Serta Pengambilan data dilakukan pada studio Lab. Kecerdasan Buatan Departemen Teknik Informatika Universitas Hasanuddin. Sedangkan jumlah daftar kata *Corpus* yang digunakan sejumlah 2000 kata dan data potongan kata (*Diphone*) berjumlah 225. Parameter unjuk kerja sistem ditinjau berdasarkan keberhasilan sistem dalam mengubah teks menjadi suara serta *Mean Opinion Score* (MOS) dari hasil kuesioner responden menggunakan 5 skenario pengujian dengan 3 kriteria penilaian, yaitu tingkat kejelasan (*Intelligibility*), kelancaran (*Fluidity*), maupun kealamian ucapan (*Naturalness*). Hasil penelitian menunjukkan tingkat keberhasilan sistem mengkonversi teks ke suara memperoleh akurasi hingga 100% pada setiap skenario untuk setiap kata yang terdapat pada database kata *Diphone*. Penelitian ini juga meninjau kualitas keluaran sistem menggunakan MOS dari 3 kriteria penilaian, menunjukkan hasil rata-rata pada setiap skenario yaitu untuk tingkat *Intelligibility* mencapai nilai 3.7, lalu tingkat *Fluidity* mencapai nilai 3.7 dan tingkat *Naturalness* mencapai nilai 3.8. Sistem *Text-to-Speech Corpus* bahasa Indonesia yang dibuat menggunakan metode *Diphone Concatenation* memiliki tingkatan kualitas hasil konversi suara yang dapat dipahami dengan ucapan yang jelas, pengucapannya lancar dari transisi antar *Diphone*, serta pengucapan dari suaranya berintonasi baik dan sesuai dengan pengucapan manusia pada umumnya.

**Kata kunci:** *Natural Language Processing, Text-To-Speech, Diphone Concatenation, Mean Opinion Score*

## ABSTRACT

The development of Natural Language Processing (NLP) in Indonesia is severely hampered due to a lack of resources such as the Corpus. This is due to the government's lack of attention to the field of language and this research requires quite a long time due to the large and varied number of words to make a Corpus. Therefore, this system was created so that it can contribute to language research and assist in the development of research in the field of NLP in Indonesia by creating a Text-To-Speech system with Diphone Concatenation method. The data were collected in the studio Lab. Artificial Intelligence Department of Informatics Engineering, University of Hasanuddin. Meanwhile the number of Corpus used is 2000 words and data of Diphone is 225. The system performance was evaluated by reviewing the system's success in convert text into sound and the Mean Opinion Score (MOS) from the results of the respondent's questionnaire. This study used 5 scenarios with 3 assessment criteria, namely range of Intelligibility, Fluidity, and the Naturalness of speech. Results showed that the highest accuracy of conversion of the text-to-speech system obtained 100% in each scenario for each word contained in the Diphone database. This research also reviews the quality of system using MOS from 3 criteria in each scenario, namely the average result for the Intelligibility value of 3.7, then Fluidity value of 3.7 and Naturalness value of 3.8. The development system conversion text to speech, show results that can be understood with clear speech, smooth pronunciation of transitions between diphones, as well as pronunciation of the voice that has good intonation and is in accordance with a generally human speech.

**Kata kunci:** *Natural Language Processing, Text-To-Speech, Diphone Concatenation, Mean Opinion Score*

## DAFTAR ISI

KATA PENGANTAR .....	i
ABSTRAK .....	iii
ABSTRACT .....	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR .....	vii
DAFTAR TABEL.....	viii
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
1.6 Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA .....	10
2.1 <i>Corpus</i> .....	10
2.2 <i>Text-To-Speech (TTS)</i> .....	13
2.3 <i>Concatenation Synthesizer</i> .....	26
2.4 Python.....	27
2.5 <i>User Interface</i> .....	30
BAB III METODOLOGI PENELITIAN .....	33
3.1 Tahapan Penelitian .....	33
3.2 Waktu dan Lokasi Penelitian.....	35
3.3 Instrumen Penelitian.....	35
3.4 Teknik Pengambilan Data .....	37
3.5 Sampel Data .....	39
3.6 Perancangan System.....	42
3.7 Analisis Kinerja Sistem.....	46



BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM.....	50
4.1 <i>User Interface System</i> .....	50
4.2 Pengujian Sistem.....	52
BAB V PENUTUP .....	62
5.1 Kesimpulan.....	62
5.2 Saran.....	63
DAFTAR PUSTAKA .....	64
LAMPIRAN .....	67

## DAFTAR GAMBAR

<b>Gambar 2. 1</b> Pemanfaatan <i>Text-to-Speech</i> .....	18
<b>Gambar 2. 2</b> Sistem <i>Text-To-Speech</i> .....	18
<b>Gambar 2. 3</b> Urutan Proses <i>Text-To-Speech</i> .....	21
<b>Gambar 2. 4</b> Besaran-besaran Dalam Setiap Tahap Proses <i>Text-To-Speech</i> .....	25
<b>Gambar 2. 5</b> Skema Proses <i>Concatenation</i> TTS .....	27
<b>Gambar 3. 1</b> Diagram Tahapan Penelitian .....	33
<b>Gambar 3. 2</b> a. Laptop Asus Vivobook, b. Mic Audio Technical, c. Samson Pop Filter, d. Hercules Stand Mic, e. Kabel XLR, f. Alctron PF36B, g. Ruangan Studio .....	37
<b>Gambar 3. 3</b> Daftar Kata <i>Corpus</i> bahasa Indonesia .....	39
<b>Gambar 3. 4</b> Ilustrasi Pengambilan Data Suara.....	40
<b>Gambar 3. 5</b> Contoh Proses Data Sampel Suara Untuk Kata “Makan” .....	41
<b>Gambar 3. 6</b> Daftar Potongan <i>Diphone</i> .....	40
<b>Gambar 3. 7</b> Contoh Proses Pemotongan Sinyal Sampel Suara Kata "Makan" .....	41
<b>Gambar 3. 8</b> Contoh Sinyal Sampel Suara <i>Diphone</i> “Ma” Yang Berasal Dari Sinyal Suara Kata “Makan”.....	41
<b>Gambar 3. 9</b> <i>Flowchart</i> Rancangan System .....	46
<b>Gambar 4. 1</b> <i>User Interface System</i> .....	50
<b>Gambar 4. 2</b> Kata yang dikonversi .....	51

## DAFTAR TABEL

<b>Tabel 3. 1</b> Kriteria Penilaian <i>Intelligibility</i> (Tingkat Kejelasan).....	48
<b>Tabel 3. 2</b> Kriteria penilaian <i>Fluidity</i> (Tingkat Kelancaran).....	49
<b>Tabel 3. 3</b> Kriteria penilaian <i>Naturalness</i> (Tingkat Kealamian).....	49
<b>Tabel 4. 1</b> Pengujian Fungsional Sistem .....	52
<b>Tabel 4. 2</b> Hasil Pengujian Fungsional Sistem.....	53
<b>Tabel 4. 3</b> Skenario 1 Pengujian Sistem (1 Kata).....	55
<b>Tabel 4. 4</b> Skenario 2 Pengujian Sistem (2 Kata).....	55
<b>Tabel 4. 5</b> Skenario 3 Pengujian Sistem (3 Kata).....	55
<b>Tabel 4. 6</b> Skenario 4 Pengujian Sistem (Kata Berimbuhan).....	56
<b>Tabel 4. 7</b> Skenario 5 Pengujian Sistem (Kalimat) .....	56

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perkembangan teknologi saat ini sangat pesat dan mempengaruhi berbagai aspek kehidupan manusia. Teknologi diciptakan sebagai alat bantu yang dapat mempermudah manusia dalam menyelesaikan suatu pekerjaan (Bachtiar, 2017). Banyaknya teknologi yang dikembangkan untuk menyelesaikan masalah yang ditemukan dalam kehidupan sehari-hari, sehingga hal ini terus mendorong peneliti untuk mengembangkan teknologi yang semakin kompleks (Novela, 2021). Di Indonesia khususnya penelitian pada bidang bahasa telah dilakukan sejak lebih dari 10 tahun yang lalu. Bidang ini merupakan bidang multi disiplin serta belum mendapat prioritas utama untuk didanai dari berbagai sumber dana penelitian, sehingga untuk kurun waktu yang lama, belum ada hasil yang dapat diaplikasikan dari penelitian di bidang ini, sehingga untuk pengembangannya masih sangat kurang. Selain itu, sebelum melakukan pengembangan sistem yang dapat diaplikasikan, diperlukan studi, pengukuran dan analisis-analisis fenomena linguistik yang selama ini hampir tidak pernah dilakukan oleh para ahli linguistik di Indonesia. Akibatnya, para pengembang teknologi bahasa di Indonesia harus melakukan penelitian linguistik dahulu untuk bahasa Indonesia sebelum memasuki tahap pengembangan sistem (A. Arman, 2004).

Ilmu linguistik modern disemarakkan dengan kemunculan suatu cabang ilmu yang relatif baru, yaitu linguistik korpus. Ilmu ini secara

khusus meneliti bahasa melalui seperangkat data yang bersifat alamiah, riil sesuai penggunaannya, baik itu data tulisan maupun data lisan yang ditranskripsikan (Svenja Adolphs, 2006). Dalam pandangan Cheng, linguistik korpus bekerja dengan menganalisis data melalui proses kompilasi berdasarkan kategori dan satuan analisisnya (Winnie Cheng, 2012). Di tahap berikutnya, kategori dan satuan analisis yang paling penting dalam lingkup linguistik korpus adalah daftar kata, kata kunci, dan konkordansi. Daftar kata mengandung sejumlah besar kata yang terdapat dalam data kebahasaan dan jumlah frekuensi kemunculannya dalam teks sumber. Sebelum ditemukan dan berkembangnya teknologi komputer, semua data kebahasaan dalam bentuk tertulis masih berupa manuskrip atau tulisan biasa di atas kertas ataupun objek lain. Sejak perkembangan teknologi dan dimulainya era digital, kini data kebahasaan yang berbentuk tulisan atau teks serta bahasa lisan yang diubah menjadi tulisan digital dapat disimpan dan diolah secara lebih mudah (Nur Hizbullah, 2016). Sehingga melalui ini linguistik korpus mengambil bagian, yaitu mengolah data kebahasaan digital.

Pada dasarnya *Corpus* merupakan suatu metode yang menggunakan sekumpulan data biasa maupun data digital dari bahan-bahan bahasa yang berasal dari penggunaan berbagai genre, ragam, dan bahan lisan maupun tertulis yang menjamin keragaman yang seluas-luasnya dan menghindari penggunaan bahasa yang sangat sempit yang terkumpul dalam suatu sumber yang disebut *Corpus* atau korpora (*bank*). Data yang terkumpul bisa dalam bentuk tertulis yang berisi informasi suatu Bahasa, mulai dari

tataran kata, struktur, makna, yang dapat digunakan untuk kepentingan penelitian. Sebelum kemunculan dan berkembangnya teknologi komputer, semua data kebahasaan dalam bentuk tertulis masih berupa manuskrip atau tulisan biasa di atas kertas ataupun objek lain. Sehingga diperlukan sistem yang dapat mengolah data yang berbentuk manuskrip atau tulisan biasa ke dalam bentuk bahasa digital/suara untuk mempermudah pengolahan data tersebut. Oleh karena permasalahan tersebut maka perlu dibuat sebuah sistem *Text-to Speech* pada *Corpus* bahasa Indonesia.

Berdasarkan perkembangan interaksi antara manusia dan komputer yang telah mengalami kemajuan dari waktu ke waktu, maka terdapat suatu sistem yang dapat melakukan dua konversi yaitu teks ke fonem dan konversi fonem ke ucapan dengan menginput suatu teks lalu akan menghasilkan output ucapan atau biasa disebut dengan *Text-to-speech* (TTS) (Salita Ulitia. P, 2016). TTS ini dapat mengubah suatu teks menjadi ucapan secara otomatis dengan cara fonetisasi (penyusunan fonem-fonem untuk membentuk ucapan). Dengan teknologi TTS, dimungkinkan sebuah perangkat elektronik mampu berkomunikasi dan berinteraksi dengan manusia tidak hanya melalui tulisan, namun juga dalam bentuk lisan menggunakan Bahasa yang digunakan sehari-hari. Namun, dalam mengkonversi teks menjadi suara untuk menghasilkan suara yang dapat dimengerti sehingga membutuhkan penyimpanan yang sangat besar untuk data fonem dan prosodi dalam Bahasa Indonesia.

Adapun beberapa penelitian sebelumnya terkait *Text-to-Speech* yang telah dilakukan, yaitu oleh Sri Wahyuni (2017) terkait dengan Normalisasi kata singkatan

sms pada sistem *Text-to-Speech* untuk penutur berbahasa Indonesia berbasis android, penelitian ini hanya melakukan normalisasi kata singkatan berdasarkan kata per kata bukan berdasarkan konteks kalimat pada teks sms. Sehingga dari hasil uji keberhasilan sistem dalam membaca kata singkatan pada teks sms hanya mencapai 67.85% dikarenakan terbatasnya kata singkatan pada database yang telah dibuat, tetapi setelah proses update database keberhasilan sistem dalam membaca kata mencapai 100% (Wahyuni, 2017). Nurpadila Muslimin (2017) melakukan penelitian terkait pembaca pesan *Text-to-Speech* berbasis android menggunakan *Framework Google API*, penelitian ini merancang sebuah sistem yang dapat membaca pesan pada aplikasi Line, Telegram dan WhatsApp. Namun dari hasil uji sistem keberhasilan membaca pesan masih cukup rendah, yaitu hanya mencapai 78.57% dari turunan *NotificationListenerService*, sedangkan ketika sistem ditambahkan kata singkatan akurasi sistem mencapai 92.85% (N. Muslimin, 2017). Selanjutnya, Arfan Al Mukaddas (2018) juga melakukan penelitian terkait penerapan *Text-To-Speech* pada sistem pembaca teks SMS berbahasa Indonesia berbasis Android. Penelitian ini merancang sebuah sistem yang dapat membaca teks SMS yang diterima oleh pengguna *Smartphone* yang di sebut *Text-To-Speech* sehingga pengguna kendaraan bermotor dapat tetap mengetahui isi pesan tanpa harus membuka *smartphone*. Penelitian ini menggunakan metode *Diphone Concatenation* dengan tingkat akurasi mencapai 97,59%. Dengan tingkat kejelasan yaitu 3, tingkat kelancaran pengucapan 2,25 dan tingkat kealamian pengucapan yaitu 1,75, ini dikarenakan sistem hanya memiliki

sedikit intonasi dan pengucapannya tidak sesuai dengan pengucapan manusia pada umumnya (A. Almukaddas, 2018).

Oleh karena itu berdasarkan beberapa permasalahan penelitian sebelumnya, maka perlu untuk dibuat sebuah sistem *Corpus* berbahasa Indonesia yang memiliki ragam intonasi yang lebih banyak dengan menambah database bahasa Indonesia, serta mengambil data dengan ucapan yang sesuai dengan pengucapan manusia pada umumnya dan membuat sistem *Text-To-Speech* dengan menggunakan metode *Diphone Concatenation* agar konversi dari teks ke suara dapat dilakukan dan menghasilkan kualitas suara dengan baik. Hal ini ditinjau berdasarkan unjuk kerja (performansi) sistem dari sisi akurasi dan *Mean Opinion Score* (MOS) untuk mengukur tingkat kejelasan, kelancaran, dan kealamian ucapan dari sistem *Text-to-Speech*. Untuk itu, maka tugas akhir yang dilakukan berjudul **“Pembuatan *Corpus* Berbahasa Indonesia Dengan Metode *Diphone Concatenation* Untuk Sistem *Text-To-Speech*”**

## 1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana membuat sistem *Text-To-Speech* dengan menggunakan metode *Diphone Concatenation*?
2. Bagaimana unjuk kerja (performansi) sistem yang dibuat berdasarkan tingkat kejelasan, kelancaran, dan kealamian ucapan dari sistem *Text-to-Speech* yang dibuat?



### **1.3 Tujuan Penelitian**

Tujuan akhir dari penelitian ini adalah:

1. Untuk membuat sistem *Text-to-Speech* dengan metode *Diphone Concatenation*.
2. Untuk mengukur unjuk kerja (performansi) sistem yang dibuat dari tingkat kejelasan, kelancaran, dan kealamian ucapan dari sistem *Text-to-Speech* yang dibuat.

### **1.4 Manfaat Penelitian**

Manfaat akhir dari penelitian ini adalah:

Dapat dijadikan referensi dan wawasan terhadap peneliti lain untuk mengembangkan penelitian selanjutnya khususnya pada bidang *Natural Language Processing* untuk sistem *Text-to-Speech*. Serta ilmu pengetahuan terkait teknologi industri 4.0, khususnya mengenai *Corpus* berbahasa Indonesia dengan sistem *Text-to-Speech*.

### **1.5 Batasan Masalah**

Adapun yang menjadi batasan masalah dalam penelitian ini adalah:

1. Sistem *Text-To-Speech* menggunakan metode *Diphone Concatenation*.
2. Data yang digunakan adalah data suara yang diambil sendiri oleh penulis di studio Lab. Kecerdasan Buatan Departemen Teknik Informatika Universitas Hasanuddin.
3. Hanya membaca sebuah kalimat dengan penggalan kata.

4. Menggunakan Bahasa pemrograman Python.
5. Pengujian dilakukan dengan menggunakan 5 skenario, yaitu skenario 1 menggunakan 1 kata, skenario 2 menggunakan 2 kata, skenario 3 menggunakan 3 kata, skenario 4 menggunakan kata imbuhan dan skenario 5 menggunakan frasa dengan susunan 8 kata.

### **1.6 Sistematika Penulisan**

Untuk memberikan gambaran umum mengenai isi tulisan secara keseluruhan, maka akan diuraikan beberapa tahapan dari penulisan secara sistematis, yaitu:

## **BAB I PENDAHULUAN**

Pada Bab I ini menguraikan terkait latar belakang, yang menjabarkan alasan dilakukannya penelitian terkait pembuatan *Corpus* Berbahasa Indonesia, berdasarkan peluang penelitian dengan uraian penelitian awal tentang *Corpus* untuk sistem *Text-to-Speech*. Pada bagian ini juga menjabarkan terkait rumusan masalah, tujuan, manfaat, ruang lingkup serta sistematika penulisan penelitian dibahas pada bagian ini.

## **BAB II TINJAUAN PUSTAKA**

Pada Bab II ini berisi penjelasan terkait landasan teori-teori yang berhubungan dengan penelitian yang dilakukan, meliputi dari *Corpus*, *Text-To-Speech*, *Natural Language Processing*, serta metode *Diphone Concatenation*. Serta diuraikan pula tentang tinjauan pustaka yang merupakan penjelasan tentang hasil-hasil penelitian sebelumnya yang berkaitan dengan penelitian yang dilakukan. Landasan teori yang

dijabarkan merupakan suatu penjelasan tentang sumber acuan terbaru dari pustaka primer seperti buku, artikel, jurnal, prosiding dan tulisan asli lainnya untuk mengetahui perkembangan penelitian yang relevan dengan judul atau topik penelitian yang dilakukan dan juga sebagai acuan dalam memecahkan masalah-masalah yang diteliti.

### **BAB III METODOLOGI PENELITIAN**

Pada Bab III ini berisi tentang penjelasan tentang tahapan-tahapan yang dilakukan di dalam penelitian ini berdasarkan *Research & Development* (R&D), serta waktu dan lokasi penelitian yang dilakukan selama  $\pm 2$  tahun di Lab Kecerdasan Buatan. Adapun terkait penjelasan jenis penelitian ini yang bersifat analisis dengan studi literature yang ada (*Library Research*) dengan didukung dari beberapa instrument penelitian. Pembahasan tentang teknik pengambilan data pada bab ini dijelaskan dengan menunjukkan ilustrasi berupa gambar proses pengambilan data. Sampel data yang digunakan juga dijelaskan pada bagian ini berupa jumlah data maupun daftar kata yang digunakan sebagai sampel dalam penelitian ini. Serta terdapat rancangan sistem yang ditunjukkan pada setiap proses dalam bentuk flowchart dari setiap tahapan yang ada. Adapun teknik evaluasi kinerja sistem yang dibuat ditunjukkan pada bagian ini dengan menunjukkan persamaan yang telah dilampirkan menggunakan evaluasi akurasi dan *Mean Opinion Score*, sehingga kinerja sistem dapat diketahui.

## **BAB IV HASIL DAN PEMBAHASAN**

Pada Bab IV ini berisi tentang penjabaran hasil penelitian berdasarkan teknik evaluasi kinerja sistem yang digunakan. Pada bagian ini juga menjelaskan terkait hasil dan pembahasan berdasarkan seberapa akurat sistem dalam mengkonversi *Text-to-Speech*. Hasil analisis kinerja sistem *Text-to-Speech* Corpus berbahasa Indonesia yang telah dibuat dirangkum dalam bentuk Tabel dan gambar berupa *User Interface* sistem *Text-to-Speech* Corpus Bahasa Indonesia. Serta pembahasan hasil penelitian dibahas pada setiap bagian pada setiap data yang diujikan berdasarkan daftar kata.

## **BAB V PENUTUP**

Pada Bab V ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan dalam penelitian ini, dengan merujuk pada rumusan masalah serta saran-saran untuk pengembangan lebih lanjut untuk menyempurnakan kekurangan-kekurangan atau capaian-capaian yang belum dicapai dalam penelitian ini. Sehingga kedepannya penelitian yang dilakukan dapat dikembangkan dan bisa memperoleh hasil yang jauh lebih baik.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 *Corpus***

*Corpus* merupakan kajian linguistik yang secara khusus meneliti bahasa melalui seperangkat data yang bersifat alamiah, riil sesuai penggunaannya, baik itu data tulisan maupun data lisan yang ditranskripsikan (Svenja Adolphs, 2006). Serta metode yang menggunakan sekumpulan data biasa maupun data digital dari bahan-bahan bahasa yang berasal dari penggunaan berbagai genre, ragam, dan bahan lisan maupun tertulis yang menjamin keragaman yang seluas-luasnya yang terkumpul dalam suatu sumber yang disebut *Corpus* atau korpora (*bank*). Data yang terkumpul bisa dalam bentuk tertulis yang berisi mengenai informasi suatu Bahasa, mulai dari tataran kata, struktur, makna, yang dapat digunakan untuk kepentingan penelitian (Nur Hizbullah, 2016).

Sebelum kemunculan dan berkembangnya teknologi komputer, semua data kebahasaan dalam bentuk tertulis masih berupa manuskrip atau tulisan biasa di atas kertas ataupun objek lain. *Corpus* muncul pada era 1960-an, sampai sekarang sudah banyak tersedia berbagai bentuk dan ragam data kebahasaan di dunia dan sejak ditemukannya teknologi komputer dan dimulainya era digital, kini data kebahasaan yang berbentuk teks atau tulisan serta Bahasa lisan yang diubah menjadi tulisan digital dapat disimpan dan diolah secara lebih mudah.

*Corpus* bukan hanya kumpulan data Bahasa namun *Corpus* juga yang mewakili beberapa bagian dari Bahasa. Sehingga pembuatan *Corpus* tergantung kepada tujuan

apa yang ingin dicapai. *Corpus* juga memiliki berbagai jenis, diantaranya yaitu (Husen, 2018):

1. *General Corpus* terdiri dari teks-teks yang umum, teks yang bukan memiliki satu jenis teks, subjek. Contohnya adalah *British National Corpus* terdiri dari 100 juta kata, *Brown Corpus* terdiri dari 1 juta kata *LOB Corpus* terdiri dari 1 juta kata.
2. *Specialized Corpus* adalah *Corpus* yang berisi teks yang dibuat dengan tujuan yang spesifik, seperti untuk memberikan sebuah topik. Contohnya, *Cambridge and Nottingham Corpus of Discourse in English (CANCODE) (informal registers of British English)* terdiri dari 5 juta kata, *Michigan Corpus of Academic Spoken English (MICASE) (spoken registers in a US academic setting)* terdiri dari 5 juta kata.
3. *Regional Corpora, Corpus* yang dibangun untuk merepresentasikan keberagaman atau variasi bahasa dari suatu daerah, seperti dialek.
4. *Monolingual Corpus* merupakan *Corpus* yang paling banyak koleksinya. *Corpus* ini dibuat hanya terdiri dari satu bahasa saja. *Corpus* ini biasanya sudah dilengkapi dengan anotasi. *Corpus* ini digunakan untuk berbagai tugas, seperti memeriksa penggunaan kata yang tepat atau mencari kombinasi kata yang paling cocok, untuk penelitian ilmiah, seperti mengidentifikasi pola yang sering atau tren baru dalam bahasa.

5. *Parallel Corpus* adalah *Corpus* yang terdiri dari dua *Corpus* monolingual. Contohnya terjemahan novel. Kedua bahasa yang ada perlu dipasangkan dengan benar. Pengguna bisa mencari contoh kata atau frasa pada satu bahasa dan hasilnya akan ditampilkan bersama dengan bahasa yang lain dengan kalimat yang sesuai.
6. *Multilingual Corpus* serupa dengan *Parallel Corpus*. Kedua istilah ini sering digunakan secara bergantian. *Multilingual Corpus* berisi teks dari beberapa bahasa yang semua terjemahan teks yang sama dan dipasangkan dengan cara yang sama seperti *Parallel Corpus*.
7. *Comparable Corpus* adalah pasangan dari dua atau lebih *Monolingual Corpus* yang mana teksnya berhubungan dengan topik yang sama. Contohnya, *International Corpus of English (ICE)* yang terdiri dari 1 juta kata.
8. *Learner Corpus* adalah *Corpus* teks yang dibuat oleh peneliti bahasa. *Corpus* ini digunakan untuk mempelajari kesalahan dan masalah peneliti saat sedang belajar bahasa asing. Contohnya, *Louvain Corpus of Native English Essays (LOCNEE)*, *International Corpus of Learner English (ICLE)* terdiri dari 20 ribu kata.
9. *Diachronic Corpus* adalah *Corpus* yang berisi teks yang dibuat dari periode yang berbeda dan digunakan untuk pengembangan. Contohnya, *Helsinki Corpus – 700 to 1700 texts*, terdiri dari 1,5 juta kata.

10. Beberapa abad terakhir, bertambah dimensi baru dalam bidang linguistic.

Bidang yang ditambahkan yaitu *Linguistik Corpus*, merupakan bagian penting dari komputasi linguistik. *Corpus* Linguistik menyediakan sejumlah basis data bahasa yang besar yang terakumulasi secara sistematis dari berbagai bidang penggunaan bahasa yang sebenarnya mengikuti beberapa metode statistik dan teknik pengambilan data sampel.

## **2.2 Text-To-Speech (TTS)**

*Text-to-Speech* (TTS) pada dasarnya adalah sebuah sistem yang dapat mengubah teks sebagai bentuk masukan menjadi keluaran dalam bentuk suara. Sistem ini dapat mengubah suatu teks menjadi ucapan secara otomatis dengan cara *fonetisasi* yaitu penyusunan fonem-fonem untuk membentuk ucapan.

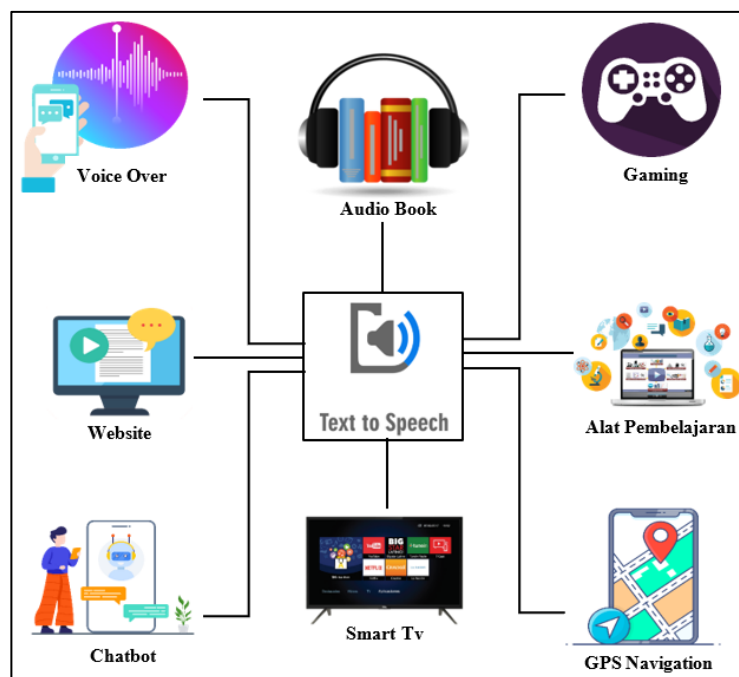
Sistem pensintesis ucapan juga biasa disebut dengan sistem tulis-ucap. Sistem ini dapat digunakan untuk berbagai aplikasi, sebagai contoh sistem informasi tagihan telepon atau sistem informasi lainnya yang diucapkan secara lisan. Metode sintesis ucapan memungkinkan mesin dapat melewati perintah atau informasi kepada pengguna lewat “ucapan”. Proses ini melibatkan pemecahan kata menjadi fonem, menganalisis penanganan khusus dari teks seperti angka, jumlah mata uang, perubahan nada suara, dan juga pemberian tanda-tanda baca (Hapsari, 2014).

Konverter ini banyak digunakan untuk mempermudah manusia yang memiliki kekurangan seperti tuna wicara dan tunanetra bahkan konverter ini juga bisa digunakan



untuk mempermudah memberi pengajaran kepada anak-anak atau orang-orang yang masih belajar tulis menulis dan membaca (Rugmiaga and Huda, 2011).

Dengan teknologi TTS, dimungkinkan sebuah perangkat elektronik mampu berkomunikasi dan berinteraksi dengan manusia tidak hanya melalui tulisan, namun juga dalam bentuk lisan menggunakan Bahasa yang di gunakan sehari-hari. Namun, dalam mengkonversi teks menjadi suara untuk menghasilkan suara yang dapat dimengerti sehingga membutuhkan penyimpanan yang sangat besar untuk data fonem dan prosodi dalam Bahasa Indonesia. Dukungan hardware dan software yang memadai memungkinkan sistem TTS diimplementasikan pada perangkat tersebut (Tritoasmoro, 2006). Adapun beberapa pemanfaatan yang dapat diterapkan dalam mengimplementasikan sistem *Text-to-Speech*, dapat dilihat pada Gambar 2.1.



**Gambar 2.1** Pemanfaatan *Text-to-Speech*

Pengimplentasian *Text-to-Speech* tentunya akan membawa berbagai macam manfaat, baik itu dari pengguna maupun untuk bisnis serta organisasi. Sehingga sistem *Text-to-Speech* dapat digunakan dalam berbagai hal, seperti (Wicara, 2022):

### 1. *Voice Over*

*Voice Over* dapat digunakan untuk para pembuat konten yang memiliki kendala dalam menemukan suara narator yang sesuai. Dengan menggunakan teknologi *Text-to-Speech* (TTS) untuk menjadi narator konten. Pengguna sistem *Text-to-Speech* hanya menginput teks yang diinginkan, maka sistem akan mengonversinya menjadi suara secara otomatis.

### 2. *Audio Book*

Audiobook adalah produk buku bersuara dengan narator sebagai pembaca keseluruhan isi buku. Suara narator yang digunakan pun bermacam-macam, bisa dari *human voice actor* atau bisa juga menggunakan teknologi *Text-to-Speech* yang pintar.

### 3. Alat pembelajaran

Berdasarkan riset di Barcelona University, penggunaan *Text-to-Speech* (TTS) dalam bidang pendidikan menunjukkan efisiensi yang cukup tinggi bagi siswa. Diperkuat lagi dengan berbagai penelitian lain yang menunjukkan bahwa pembelajaran melalui teks dan audio dinilai lebih efektif. Teknologi TTS dapat membantu siswa untuk meningkatkan kemampuan membaca, mengenali kata,

mengulang informasi, meningkatkan kecepatan membaca, mempermudah mengingat informasi, lebih fokus, dan terlebih lagi dapat membantu siswa penyandang disabilitas atau disleksia agar mudah menerima pelajaran.

#### 4. *Website*

*Text-to-Speech* (TTS) juga bisa dimanfaatkan untuk website dengan mengkonversi konten website menjadi suara. Ketika user berkunjung ke website, mereka tidak hanya disambut oleh tulisan-tulisan saja, melainkan juga oleh suara dari TTS yang siap memberikan informasi penting dari website.

#### 5. *Customer Service*

Banyak perusahaan yang menggunakan *chatbot* sebagai upaya meningkatkan kepuasan pelanggan. Chatbot juga bisa diintegrasikan dengan teknologi *Text-to-Speech*. Sehingga kepuasan pelanggan akan meningkat begitu pesat bila dilayani oleh *chatbot* yang telah dipadukan dengan *teknologi Text-to-Speech*. Dengan perpaduan kedua sistem ini, maka *Customer Service* tidak perlu membaca, karena segala informasi mampu disajikan dalam bentuk suara. Mulai dari informasi perusahaan, produk, layanan hingga berbagai kebutuhan lain.

#### 6. Pemberitahuan di Tempat Umum

Pemberitahuan di tempat umum bisa menggunakan suara rekaman dari *Voice Over Talent*, akan tetapi bisa juga menggunakan suara dari teknologi *Text-to-Speech* (TTS). Bukan hanya menawarkan hasil suara yang natural, jernih, stabil,

TTS voice over juga menawarkan waktu produksi yang singkat dan mudah. Hal ini tentu akan menguntungkan perusahaan travel dan transportasi.

### 7. *Gaming*

TTS dimanfaatkan oleh para pelaku di dunia *entertainment*, salah satunya adalah dunia *gaming*. Misalnya, penggunaan suara untuk digital avatar yang dapat bicara di dalam *Video Game* dan *Virtual Reality (VR)*. Selain itu, TTS juga dimanfaatkan untuk menyuarakan narasi pada games, tips, petunjuk atau *storytelling* lainnya.

### 8. *Smart TV*

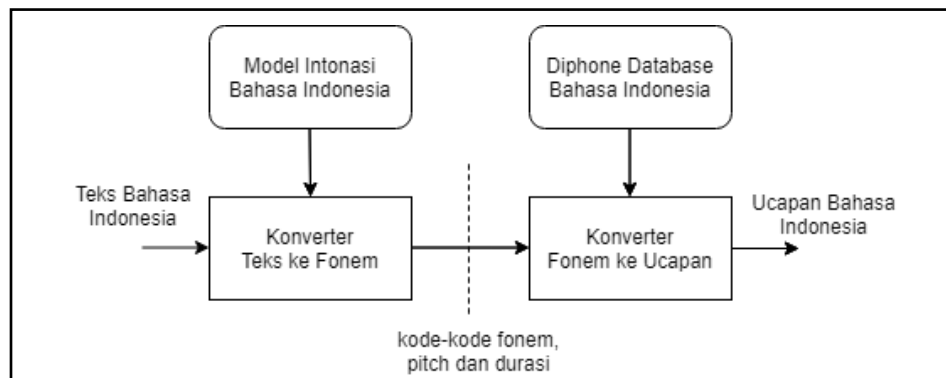
Saat ini, teknologi *Smart TV* memungkinkan pemiliknya mengakses internet atau berbagai channel televisi di penjuru dunia dengan mudah. Mengimplementasikan TTS ke dalam sistem *Smart TV* akan membuat pengguna dapat menikmati berbagai berita, laporan cuaca, berita olahraga, berbagai *website*, blog, majalah online, atau deskripsi dan review film, dalam bentuk suara.

### 9. GPS dan Sistem Navigasi

Rupanya bidang otomotif juga dapat memanfaatkan teknologi *Text-to-Speech* dalam sistem mereka, misalnya GPS dan sistem navigasi. Untuk memandu pengemudi menuju sebuah tempat, pengguna sistem navigasi membutuhkan suara pemandu yang jelas. Dengan demikian, pengemudi akan merasa nyaman berinteraksi dengan GPS atau sistem navigasi didalam kendaraannya.

Dari beberapa pemanfaatan yang ada tentunya digunakan untuk mempermudah dan solusi terhadap apa yang ingin digunakan dengan sistem *Text-to-Speech* yang dirancang. Pada sistem *Text-To-Speech* prinsipnya terdiri dari dua sub sistem, yaitu:

1. Konversi teks ke fonem ini berfungsi untuk mengolah kalimat masukan dalam suatu bahasa tertentu yang berbentuk teks menjadi urutan kode-kode bunyi yang direpresentasikan dengan kode fonem, durasi serta nadanya. Kode-kode fonem adalah kode yang merepresentasikan unit bunyi yang ingin diucapkan. Pengucapan kata atau kalimat pada prinsipnya adalah urutan bunyi atau secara simbolik adalah urutan kode fonem.
2. Kemudian konversi fonem ke ucapan yang akan menerima masukan berupa kode-kode fonem serta nada dan durasi yang dihasilkan oleh konversi sebelumnya. Berdasarkan kode-kode tersebut, bagian konversi fonem ke ucapan maka akan menghasilkan bunyi atau sinyal ucapan dengan kalimat yang ingin diucapkan.



**Gambar 2. 2** Sistem *Text-To-Speech* (Hapsari, 2014)

Ada beberapa alternatif Teknik yang dapat digunakan untuk implementasi bagian ini. Dua teknik yang banyak digunakan adalah *formant synthesizer*, serta *Diphone Concatenation*.

1. *Formant synthesizer* bekerja berdasarkan suatu model matematis yang akan melakukan komputasi untuk menghasilkan sinyal ucapan yang diinginkan. *Synthesizer* jenis ini telah lama digunakan pada berbagai aplikasi. Walaupun dapat menghasilkan ucapan dengan tingkat kemudahan interpretasi yang baik, *synthesizer* ini tidak dapat menghasilkan ucapan dengan tingkat kealamian yang tinggi.

2. *Synthesizer* yang menggunakan teknik *Diphone Concatenation* bekerja dengan cara menggabung-gabungkan segmen-segmen bunyi yang telah direkam sebelumnya. Setiap segmen berupa *Diphone* (gabungan dua buah fonem). *Synthesizer* jenis ini dapat menghasilkan bunyi ucapan dengan tingkat kealamian (*Naturalness*) yang tinggi.

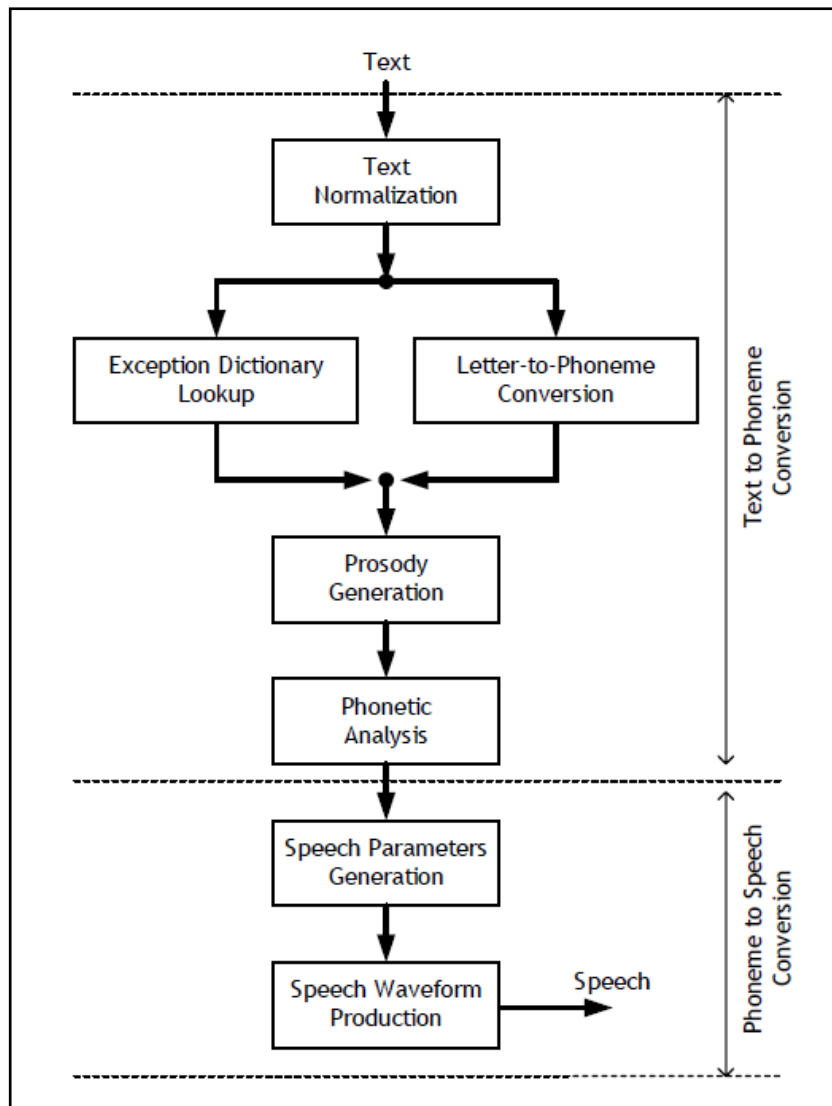
Struktur sistem seperti di atas pada prinsipnya merupakan konfigurasi tipikal yang digunakan pada berbagai sistem *Text-to-Speech* berbagai bahasa. Namun demikian, pada setiap sub-sistem terdapat sifat-sifat serta proses-proses yang sangat spesifik dan sangat tergantung dari bahasanya.

Konversi dari teks ke fonem sangat dipengaruhi oleh aturan-aturan yang berlaku dalam suatu bahasa. Pada prinsipnya proses ini melakukan konversi dari simbol-simbol tekstual menjadi simbol-simbol fonetik yang merepresentasikan unit bunyi terkecil dalam suatu bahasa. Setiap bahasa memiliki aturan cara pembacaan dan cara

pengucapan teks yang sangat spesifik. Hal ini menyebabkan implementasi unit konverter teks ke fonem menjadi sangat spesifik terhadap suatu bahasa.

Untuk mendapatkan ucapan yang lebih alami, ucapan yang dihasilkan harus memiliki intonasi (*prosody*). Secara kuantisasi, prosodi adalah perubahan nilai *pitch* (frekuensi dasar) selama pengucapan kalimat dilakukan atau *pitch* sebagai fungsi waktu. Pada prakteknya, informasi pembentuk prosodi berupa data-data *pitch* serta durasi pengucapannya untuk setiap fonem yang dibangkitkan. Nilai-nilai yang dihasilkan diperoleh dari suatu model prosodi. Prosodi bersifat sangat spesifik untuk setiap bahasa, sehingga model yang diperlukan untuk membangkitkan data-data prosodi menjadi sangat spesifik juga untuk suatu bahasa. Beberapa model umum prosodi pernah dikembangkan, tetapi untuk digunakan pada suatu bahasa masih perlu banyak penyesuaian yang harus dilakukan.

Tahapan-tahapan utama konversi dari teks menjadi ucapan dapat dinyatakan dengan diagram seperti terlihat pada Gambar 2.2.



**Gambar 2. 3** Urutan Proses *Text-To-Speech*

Pada bagian normalisasi teks, terjadi pengkonversian dari input yang berupa text menjadi *Diphone* (gabungan dua buah fonem). Ketika input yang berupa teks, akronim (singkatan) ataupun angka maka bagian ini akan mengkonversikan menjadi diphone yang telah tersedia di database *Diphone*.



Pada proses pembangkitan prosody sangat memperhatikan karakter sinyal ucapan manusia, untuk mendapatkan ucapan yang lebih alami. Secara kuantisasi, prosodi adalah perubahan nilai pitch (frekuensi dasar) selama pengucapan kalimat dilakukan atau pitch sebagai fungsi waktu. Pembangkitan prosody ini bertujuan untuk memperhalus hasil proses *concatenation*. Jadi proses penggabungan *Diphone-Diphone* ini bisa menghasilkan suara yang mendekati naturalnya.

Gejala intonasi atau prosodi, mempunyai hubungan yang erat dengan struktur kalimat dan interelasi kalimat dalam suatu wacana. Karena sifatnya yang sangat subyektif, maka pembahasan intonasi lebih ditekankan pada aspek gramatikal dibandingkan aspek emosional. Sebuah ucapan dilafalkan oleh penuturnya dalam pola melodi tertentu yang diterima pendengarnya sebagai sebuah deret nada dengan tinggi yang berbeda. Tingkat tinggi nada tidak dapat ditentukan secara pasti, sangat bervariasi tergantung dari jenis kelamin, emosi atau sikap (Narullita, 2011) .

Tahap berikutnya adalah melakukan konversi dari teks yang sudah secara lengkap merepresentasikan kalimat yang ingin diucapkan menjadi kode-kode fonem. Konversi teks menjadi fonem biasanya dilakukan dengan dua cara. Sebagian proses konversi dapat dilakukan dengan aturan konversi yang sederhana dan berlaku umum untuk berbagai kondisi. Sebagian proses lainnya bersifat kondisional, tergantung dari huruf-huruf atau fonem-fonem tetangganya, bahkan terdapat bentuk-bentuk translasi yang tidak dapat ditemukan keteraturannya.

Konversi yang teratur dapat diimplementasikan dengan Tabel konversi yang berisi pasangan antara urutan huruf dan urutan fonem, bahkan mungkin hanya berisi satu huruf dan satu fonem. Aturan yang lebih sulit biasanya diimplementasikan dengan Tabel konversi yang akan diterapkan jika kondisi rangkaian huruf tetangga kiri dan kanannya terpenuhi. Contoh bentuk aturan konversi huruf ke fonem yang memenuhi teknik tersebut adalah:

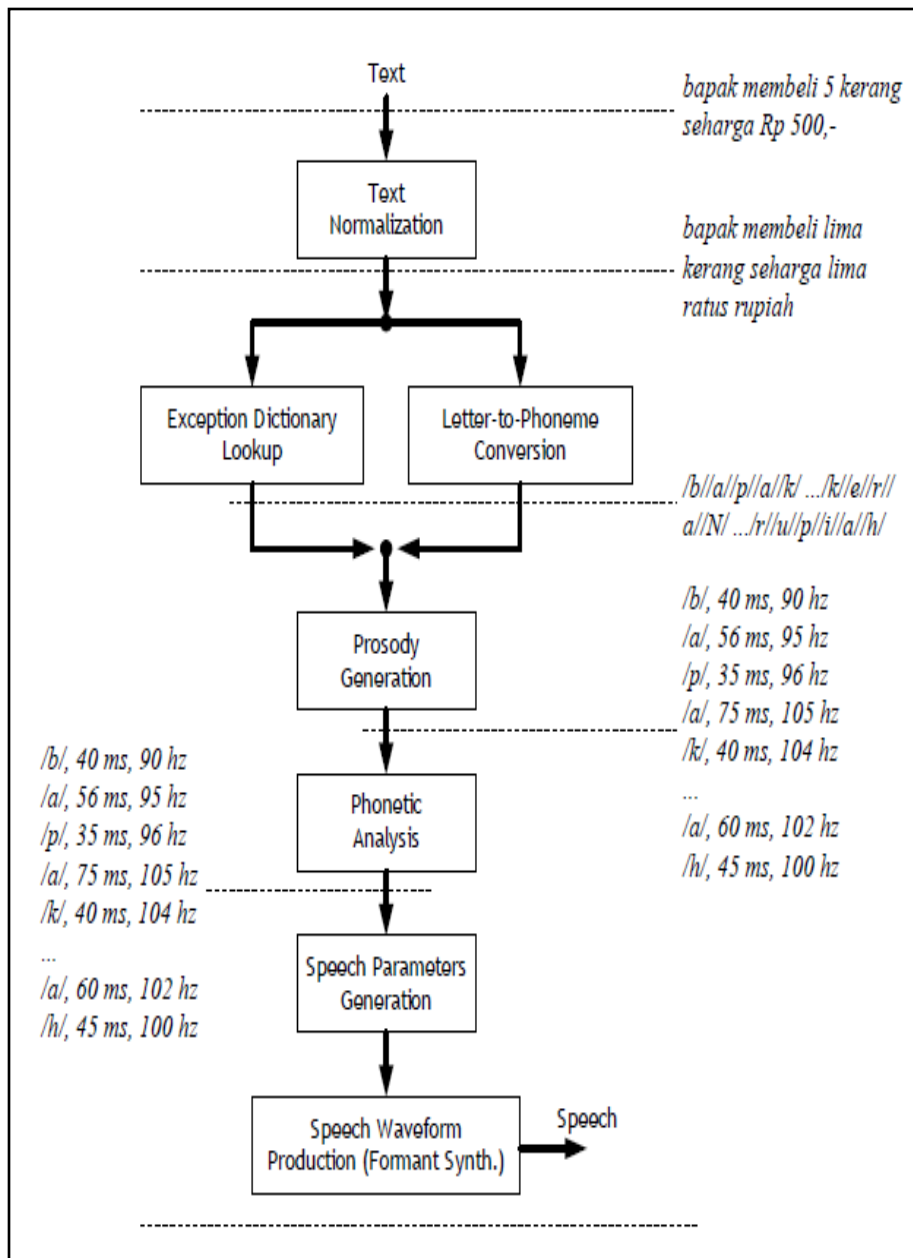
***Left-context [letter-set] right-context = phoneme string***

Huruf tertentu yang ditunjuk dalam posisi *[letter-set]* akan dikonversikan menjadi suatu fonem dalam “*phoneme string*” jika *left-context* dan *right context* terpenuhi.

Bahasa Inggris termasuk bahasa yang mempunyai keteraturan yang rendah untuk proses konversi teks ke fonem. Suatu TTS bahasa Inggris biasanya dilengkapi dengan suatu basis data yang berisi ribuan kata serta konversi padanan urutan fonemnya. Bahasa Indonesia termasuk bahasa yang jelas aturan konversinya. Sebagian besar kata dalam Bahasa Indonesia dapat dikonversikan menjadi fonem dengan aturan yang jelas dan sederhana, walaupun tetap ada kondisi-kondisi yang tidak dapat ditemukan keteraturannya. Sebagai contoh, simbol huruf e dapat diucapkan sebagai e pepet atau e taling, artinya harus dikonversikan menjadi fonem yang berbeda untuk kondisi yang berbeda. Dalam blok diagram pada Gambar 2.2, kondisi yang masih dapat ditangani oleh aturan diimplementasikan dengan blok *Letter to Phoneme Conversion*. Konversi yang tidak teratur ditangani oleh bagian *Exception Dictionary Lookup*.

Hasil dari tahap tersebut adalah rangkaian fonem yang merepresentasikan bunyi kalimat yang ingin diucapkan. Bagian *prosody generator* akan melengkapi setiap unit fonem yang dihasilkan dengan data durasi pengucapannya serta *pitch*nya. Data durasi serta *pitch* diperoleh berdasarkan kombinasi antara Tabel atau database serta model prosodi. Secara simbolik, hasil dari bagian ini sudah menghasilkan informasi yang cukup untuk menghasilkan ucapan yang diinginkan.

Satu tahap berikutnya yang masih sering dilakukan adalah *Phonetic Analysis*. Tahap ini dapat dikatakan sebagai tahap penyempurnaan, yaitu melakukan perbaikan di tingkat bunyi. Sebagai contoh, dalam bahasa Indonesia, fonem /k/ dalam kata bapak tidak pernah diucapkan secara tegas, atau adanya sisipan fonem /y/ dalam pengucapan kata alamiah antara fonem /i/ dan /a/ (Narullita, 2011).



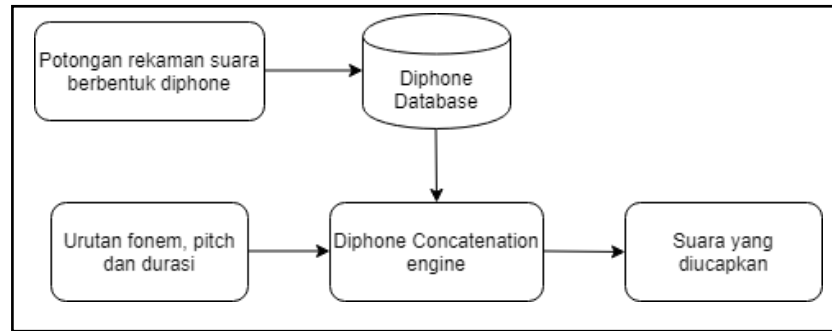
**Gambar 2. 4** Besaran-besaran Dalam Setiap Tahap Proses *Text-To-Speech*

### 2.3 Concatenation Synthesizer

*Concatenation Synthesizer* merupakan *synthesizer* yang mampu memproduksi sinyal ucapan secara otomatis melalui transkripsi grafem-ke-fonem untuk kalimat yang diucapkan. Penelitian *Concatenation Synthesizer* dalam berbagai versi bahasa sedang dikembangkan, dan dalam bahasa tertentu telah mencapai hasil yang cukup memuaskan. Namun demikian beberapa permasalahan dalam concatenation synthesizer masih belum terpecahkan secara tuntas. Pendekatan dari setiap permasalahan ini akan selalu mengacu pada tercapainya *Concatenation Synthesizer* yang memenuhi kualitas terdengar jelas (*Intelligibility*) dan kealamian (*Naturalness*).

*Concatenation* yaitu penggabung-gabungan segmen-segmen bunyi yang telah direkam sebelumnya. Setiap segmen berupa *Diphone* (gabungan dua buah fonem). Pada perekaman suara dilakukan beberapa kali agar mendapatkan hasil yang akurat.

Umumnya metode ini memang menghasilkan output sintesa ucapan yang mendekati ucapan manusia (*natural*). Namun dari aspek sejauh mana output ucapan tersebut dapat dimengerti oleh manusia (*Intelligibility*) metode ini masih memiliki kekurangan. Proses metode ini dilakukan dengan cara menggabungkan (*Concatenation*) unit-unit ucapan yang telah direkam sebelumnya. Rekaman dari segmen-segmen ucapan tersebut dapat disimpan di dalam database dalam bentuk gelombang (*uncoded*) atau dalam bentuk *encoded* disesuaikan dengan pengkodean ucapan (*speed coding method*) yang dipakai.



**Gambar 2.5** Skema Proses *Concatenation* TTS (Sudirman Melangi, 2018)

Gambar di atas menampilkan cara kerja sintesis *concatenation* berdasarkan pada rentetan bagian ucapan yang telah direkam. Dasar perangkaian sintesis adalah dengan menggabungkan segmen-segmen dari gelombang ucapan alami yang disimpan dalam database. Segmen-segmen tersebut dapat berupa kata-kata (*words*), unit subkata (*subword unit*) seperti fonem (*phonemes*), diphones dan suku kata (*syllables*). Sintesis ucapan perangkaian banyak digunakan luas, bekerja dengan prinsip pada pemilihan unit (*unit selection*). Sistem sintesis perangkaian pemilihan unit yang populer adalah *unisyn*, *clunits* dan *multisyn* (Sudirman Melangi, 2018).

## 2.4 Python

Bahasa pemrograman Python dibuat oleh Guido van Rossum pada tahun 1990 di Belanda sebagai pengganti bahasa pemrograman yang disebut ABC. Python merupakan perkembangan yang berasal dari kontribusi dari berbagai sumber dari seluruh dunia dalam bidang pemrograman bahasa meski pada akhirnya Guido lah yang mematenkannya. Pada bahasa pemrograman Python, semua orang dapat mengembangkan bahasa pemrogramannya sendiri hal ini karena Python

sendiri mempunyai sifat *Open Source* atau dapat dikatakan bahwa bahasa pemrograman Python ini gratis, dapat digunakan tanpa lisensi, dan dapat dikembangkan semampu yang dapat dilakukan.

Python adalah sebuah bahasa pemrograman dinamis yang sering digunakan dalam pengembangan aplikasi pada berbagai domain. Salah satu fitur yang tersedia pada python adalah sebagai pemrograman dinamis yang dilengkapi skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa skrip (Darmawan and Suwastono, 2015). Bahasa Python mudah dimengerti (serupa dengan bahasa Inggris), dan tidak perlu compiling, maka dari itu kita dapat menjalankan script Python begitu kita menyimpannya tanpa memerlukan langkah lanjutan.

Penulisan sintaks yang lebih fleksibel dibandingkan dengan bahasa pemrograman lainnya membuat Python menjadi mudah dipelajari. Selain itu, bahasa pemrograman Python ini memiliki efisiensi tinggi untuk struktur data level tinggi, pemrograman berorientasi objek lebih sederhana tetapi efektif, dapat bekerja pada multi platform, dan dapat digabungkan dengan bahasa pemrograman lain untuk menghasilkan aplikasi yang diinginkan.

Bahasa pemrograman Python adalah bahasa pemrograman yang mudah dibaca dan terstruktur, hal ini karena di gunakannya sistem indentasi. Yaitu memisahkan blok-blok program susunan indentasi. Jadi untuk memasukan sub-sub program dalam suatu blok, sub-sub program tersebut diletakkan satu atau lebih spasi dari kolom suatu blok

program. Menurut (Saragih, 2016) Python memiliki beberapa keunggulan dan beberapa kekurangan. Keunggulan dari Python meliputi:

1. Python mudah digunakan

Python merupakan bahasa pemrograman tingkat tinggi, artinya bahasa Python lebih mendekati bahasa manusia daripada bahasa mesin. Python juga memiliki struktur dasar yang mengharuskan pengguna untuk menginclude sistemnya sebelum membuat program. Perintah-perintah yang digunakan di Python menggunakan bahasa Inggris yang digunakan sehari-hari seperti print, input, export, dll.

2. Kompabilitas dan Kemampuan Tinggi

Python memiliki kemampuan yang tinggi dan mampu menciptakan aplikasi yang sederhana hingga aplikasi yang kompleks. Bahasa pemrograman Python mendukung pemrograman berbasis grafis (GUI Programming). Python juga memiliki kelebihan berupa pengalokasian memori secara dinamis.

3. Platform Independent

Maksudnya yaitu program yang dibuat bisa dijalankan di sistem operasi apa saja selama di sistem operasi tersebut terdapat platform Python (Interpreter Python).



#### 4. *Open Source*

Bahasa pemrograman Python bersifat gratis dan anda bisa mengembangkannya secara pribadi ataupun bersama teman-teman kalian dengan membentuk tim.

Sedangkan kekurangan dari Python yaitu:

1. Beberapa penugasan terdapat di luar jangkauan kemampuan Python sehingga Python tidak secepat atau efisien sebagai statis.
2. Akibat interpreter pada Python, maka Python bukan merupakan perangkat bantu terbaik untuk pengantar komponen kinerja kritis.
3. Untuk beberapa komponen, Python tidak dapat digunakan sebagai dasar bahasa pemrograman implementasi namun masih bisa bekerja sebagai bagian depan script interface.

Pada Python ada beberapa perintah dasar untuk menyelesaikan persamaan-persamaan matematika. Perintah-perintah tersebut bersifat sederhana, namun jika dikembangkan bisa menjadi pemrograman sebagai alat bantu hitung matematika yang rumit sekalipun. Seperti *if*, *else*, *elif (else if)*, *range*, *append*, *len*, *print* dan lain-lain

### **2.5 *User Interface***

*User Interface* adalah cara program dan pengguna berkomunikasi. Tujuannya yaitu untuk menerjemahkan informasi antara pengguna dengan sistem. *User Interface* umumnya banyak ditemukan pada *software*, *hardware*, dan sistem operasi *smartphone*. *User Interface* pada software menggunakan sistem *Graphic*

*Processing Unit* (GPU) untuk menerima perintah dari pengguna berdasarkan tampilan grafis tertentu menggunakan mouse dan keyboard. Contohnya, terjadi pada 2 sistem operasi yang umum seperti Windows dan Macintosh. *User Interface* pada hardware cenderung lebih sederhana dan lebih dominan menggunakan symbol sehingga lebih mudah dimengerti pengguna. Contohnya, remote TV yang terdapat tombol volume, keypad numerik, dan tombol-tombol fungsi lainnya sebagai symbol fungsi. Semakin maju perkembangan teknologi, perusahaan pembuat perangkat canggih berlomba-lomba membuat *User Interface* yang sederhana dan mudah dipahami (Waralalo, 2019).

*User Interface* fokus pada interaksi pengguna dengan tampilan yang muncul pada layar monitor komputer seperti elemen yang ada pada suatu program atau perangkat lunak supaya pengguna dapat mengerjakan tugas yang mereka ingin lakukan. Kesulitan atau kemudahan seorang pengguna memahami sebuah program komputer menunjukkan kualitas dari *User Interface* program tersebut. *User Interface* yang baik adalah user interface yang dapat meningkatkan pengalaman pengguna sehingga pengguna dapat memahami suatu program secara cepat dan mudah (Harris, 2017). Terdapat dua jenis *User Interface*, yaitu *Command Line Interface* (CLI) dan *Graphical User Interface* (GUI):

1. *Command Line Interface* (CLI) adalah tipe antarmuka dimana pengguna berinteraksi dengan sistem operasi melalui text-terminal. Pengguna

menjalankan perintah dan program di sistem operasi tersebut dengan cara mengetikkan baris-baris tertentu.

2. *Graphical User Interface* (GUI) adalah tipe antarmuka yang digunakan oleh pengguna untuk berinteraksi dengan sistem operasi melalui gambar-gambar grafik, ikon, menu, dan menggunakan perangkat penunjuk (*pointing device*) seperti mouse atau *track ball*. Elemen-elemen utama dari GUI bisa diringkas dalam konsep WIMP (*window, icon, menu, pointing device*).