

**SKRIPSI**

**IMPLEMENTASI ALGORITMA *KNUTH-MORRIS-PRATT*  
DALAM SISTEM PEMBUATAN KORPUS BAHASA  
INDONESIA DAN INTEGRASI KATEGORISASI FONEM  
UNTUK MENINGKATKAN AKURASI *TEXT-TO-SPEECH***

**Disusun dan diajukan oleh:**

**ARIF PUTERA WIJAYA  
D121 19 1065**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2023**

**LEMBAR PENGESAHAN SKRIPSI**

**IMPLEMENTASI ALGORITMA *KNUTH-MORRIS-PRATT* DALAM  
SISTEM PEMBUATAN KORPUS BAHASA INDONESIA DAN  
INTEGRASI KATEGORISASI FONEM UNTUK MENINGKATKAN  
AKURASI *TEXT-TO-SPEECH***

Disusun dan diajukan oleh

**ARIF PUTERA WIJAYA**  
**D121191065**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian  
Studi Program Sarjana Program Studi Teknik Informatika  
Fakultas Teknik Universitas Hasanuddin  
Pada tanggal 20 Desember 2023  
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Prof. Dr. Ir. Indrabayu,  
S.T., M.T., M.Bus.Sys., IPM,ASEAN. Eng.  
NIP 197507162002121004

Elly Warni, S.T., M.T.  
NIP 198202162008122001



Ketua Program Studi,

Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM,ASEAN. Eng.  
NIP 197507162002121004

## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini;  
Nama : Arif Putera Wijaya  
NIM : D121191065  
Program Studi : Teknik Informatika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Implementasi Algoritma *Knuth-morris-pratt* Dalam Sistem Pembuatan Korpus bahasa Indonesia Dan Integrasi Kategorisasi Fonem Untuk Meningkatkan Akurasi *Text-to-speech*

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

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

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

Gowa, 27 Desember 2023

Yang Menyatakan



Arif Putera Wijaya

## ABSTRAK

**Arif Putera Wijaya.** *Implementasi Algoritma Knuth-morris-pratt Dalam Sistem Pembuatan Korpus bahasa Indonesia Dan Integrasi Kategorisasi Fonem Untuk Meningkatkan Akurasi Text-to-speech* (dibimbing oleh Indrabayu dan Elly Warni)

Perkembangan teknologi *text-to-speech* (TTS) menjadi salah satu fokus perkembangan kecerdasan buatan dalam bidang *natural language processing* (NLP). Diantara kendala yang masih menjadi problematik adalah ketidakadaan corpus fonem bahasa Indonesia yang mengikuti kaidah bahasa, hal ini menyebabkan hambatan dalam perkembangan TTS bahasa Indonesia.

Oleh karena itu, diperlukan sistem pembuatan korpus fonem bahasa Indonesia secara otomatis yang berkualitas dengan beracuan pada kaidah linguistik bahasa Indonesia.

Penelitian ini menggunakan metode algoritma *knuth-morris-pratt* yang kemudian dihubungkan dengan kategorisasi fonem untuk menghasilkan sistem pembuatan korpus fonem yang lebih baik.

Hasil pengujian menunjukkan tingkat keberhasilan sistem dalam mengkonversi teks menjadi suara memperoleh akurasi 100% pada keseluruhan skenario uji, hal ini menunjukkan keberhasilan sistem kategorisasi fonem dalam menambahkan korpus yang digunakan dalam konversi teks menjadi suara. Penelitian ini juga meninjau kualitas keluaran sistem menggunakan *mean opinion score* (MOS) dan membandingkannya dengan kualitas keluaran sistem yang menggunakan korpus *diphone concatenation*. Hasil perbandingan ini dilihat dari tingkat kejelasan (*intelligibility*), kelancaran (*fluidity*), maupun kealamian ucapan (*naturalness*) korpus yang terintegrasi kategorisasi fonem lebih unggul dengan nilai tingkat kejelasan 3,55, tingkat kelancaran 3,58 dan tingkat kealamian 3,42, sedangkan korpus *diphone concatenation* memiliki tingkat kejelasan 2,87, tingkat kelancaran 2,9 dan tingkat kealamian 2,82. Selain itu, hasil keluaran audio sistem TTS yang menggunakan korpus terintegrasi kategorisasi fonem memiliki ukuran file yang lebih ringan 48% jika dibandingkan dengan TTS yang menggunakan korpus *diphone concatenation*. Oleh karena itu, dapat diputuskan sistem TTS dengan algoritma *knuth-morris-pratt* dan korpus terintegrasi kategorisasi fonem memiliki tingkat keberhasilan konversi dengan sangat baik dan memiliki kualitas keluaran suara yang dapat dipahami dan perangkaian ucapannya cukup jelas, pengucapan yang dihasilkan lancar, transisi antar fonem nyaman, serta memiliki tingkat intonasi yang cukup baik, yang sesuai dengan pengucapan manusia pada umumnya.

**Kata Kunci:** *Text-To-Speech*, Korpus bahasa Indonesia, Algoritma *Knuth-Morris-Pratt*, *Natural Language Processing*.

## ABSTRACT

**Arif Putera Wijaya.** *Implementation Of The Knuth-Morris-Pratt Algorithm In The Indonesian Language Corpus Creation System And Integration Of Phoneme Categorization To Increase Text-To-Speech Accuracy* (supervised by Indrabayu and Elly Warni)

The development of text-to-speech (TTS) technology is one of the focuses of artificial intelligence development in the field of natural language processing (NLP). Among the obstacles that are still problematic is the absence of a corpus of Indonesian phonemes that follow the rules of the language, this causes obstacles in the development of Indonesian TTS.

Therefore, a system of automatic generation of Indonesian phoneme corpus with quality is needed with reference to Indonesian linguistic rules.

This research uses the knuth-morris-pratt algorithm method which is then connected with phoneme categorization to produce a better phoneme corpus generation system.

The test results show the success rate of the system in converting text into sound obtaining 100% accuracy on the entire test scenario, this shows the success of the phoneme categorization system in adding the corpus used in converting text into sound. This study also reviewed the quality of system output using mean opinion score (MOS) and compared it with the quality of system output using diphone concatenation corpus. The results of this comparison are seen from the level of intelligibility, fluidity, and naturalness of speech. The corpus integrated with phoneme categorization is superior with a value of 3.55 intelligibility level, 3.58 fluency level and 3.42 naturalness level, while the diphone concatenation corpus has a 2.87 intelligibility level, 2.9 fluency level and 2.82 naturalness level. In addition, the audio output of the TTS system using the phoneme categorization integrated corpus has a file size that is 48% lighter than the TTS using the diphone concatenation corpus. Therefore, it can be decided that the TTS system with the knuth-morris-pratt algorithm and the integrated corpus of phoneme categorization has a very good conversion success rate and has a comprehensible sound output quality and the speech string is quite clear, the resulting pronunciation is smooth, the transition between phonemes is comfortable, and has a fairly good intonation level, which is in accordance with human pronunciation in general.

**Keywords:** Text-To-Speech, Indonesian Language Corpus, Knuth-Morris-Pratt Algorithm, Natural Language Processing.

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK .....	iii
ABSTRACT .....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR .....	vi
DAFTAR TABEL.....	vii
DAFTAR LAMPIRAN .....	viii
KATA PENGANTAR .....	ix
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat Penelitian .....	3
1.5 Ruang Lingkup.....	3
BAB 2 TINJAUAN PUSTAKA .....	4
2.1 <i>Text-to-speech</i> .....	4
2.2 Korpus.....	5
2.3 Algoritma <i>knuth-morris-pratt</i> .....	7
2.4 Kaidah Struktur Linguistik.....	9
2.5 <i>Concatenation Synthesizer</i> .....	16
2.6 Metode Evaluasi <i>Black box</i> dan <i>Mean Opinion Score</i> .....	18
BAB 3 METODE PENELITIAN.....	19
3.1 Waktu dan Lokasi Penelitian .....	19
3.2 Instrumen Penelitian .....	19
3.3 Tahapan Penelitian.....	19
3.4 Perancangan Sistem .....	21
3.5 Teknik Pengambilan Data.....	29
3.6 Analisis Kinerja Sistem.....	30
BAB 4 HASIL dan PEMBAHASAN .....	32
4.1 Hasil Penelitian .....	32
4.2 Keluaran Sistem .....	32
4.3 Pengujian Sistem.....	37
BAB 5 KESIMPULAN dan SARAN .....	51
5.1 Kesimpulan .....	51
5.2 Saran.....	52
DAFTAR PUSTAKA .....	53

## DAFTAR GAMBAR

Gambar 1. Sistem <i>text-to-speech</i> .....	5
Gambar 2. <i>Flowchart</i> algoritma <i>knuth-morris-pratt</i> .....	8
Gambar 3. Diftong menurun .....	13
Gambar 4. Diftong menaik.....	13
Gambar 5. Proses <i>concatenation text-to-speech</i> .....	17
Gambar 6. Tahapan penelitian .....	20
Gambar 7. Rancangan sistem.....	21
Gambar 8. <i>Flowchart</i> kategorisasi fonem.....	23
Gambar 9. <i>Flowchart</i> program fonem awal.....	26
Gambar 10. <i>Flowchart</i> program fonem tengah.....	27
Gambar 11. <i>Flowchart</i> program fonem akhir .....	28
Gambar 12. Daftar pengucapan fonem .....	29
Gambar 13. Ilustrasi pengambilan data suara .....	30
Gambar 14. Ilustrasi contoh pengambilan data suara untuk fonem “E” .....	30
Gambar 15. <i>Database</i> fonem .....	32
Gambar 16. <i>User interface system</i> .....	34
Gambar 17. Kata yang berhasil dikonversi .....	35
Gambar 18. Kata yang tidak berhasil dikonversi .....	35
Gambar 19. Notifikasi konfirmasi.....	36
Gambar 20. Kata berhasil ditambahkan .....	36
Gambar 21. Text berhasil dibersihkan .....	36

## DAFTAR TABEL

Tabel 1. Huruf vokal .....	11
Tabel 2. Huruf konsonan.....	11
Tabel 3. Konsonan gabungan.....	13
Tabel 4. Kata berimbuhan .....	14
Tabel 5. Struktur suku kata bahasa indonesia .....	15
Tabel 6. Kriteria penilaian tingkat kejelasan ( <i>Intelligibility</i> ).....	31
Tabel 7. Kriteria penilaian tingkat kelancaran ( <i>Fluidity</i> ).....	31
Tabel 8. Kriteria penilaian tingkat kealamian ( <i>Naturalness</i> ).....	31
Tabel 9. Output pasca preprocessing .....	33
Tabel 10. Output pasca kategorisasi fonem .....	33
Tabel 11. Pengujian fungsional sistem.....	37
Tabel 12. Hasil pengujian fungsional sistem .....	38
Tabel 13. Hasil pengujian skenario 1 kaegorisasi fonem.....	40
Tabel 14. Rentang nilai skenario 1 kategorisasi fonem .....	41
Tabel 15. Hasil pengujian skenario 2 kaegorisasi fonem.....	41
Tabel 16. Rentang nilai skenario 2 kategorisasi fonem .....	42
Tabel 17. Hasil pengujian skenario 3 kaegorisasi fonem.....	43
Tabel 18. Rentang nilai skenario 3 kategorisasi fonem .....	43
Tabel 19. Hasil pengujian skenario 4 kaegorisasi fonem.....	44
Tabel 20. Rentang nilai skenario 4 kategorisasi fonem .....	45
Tabel 21. Perbandingan hasil pengujian kategorisasi fonem dan <i>diphone</i> <i>concatenation</i> .....	46
Tabel 22. Ringkasan hasil penelitian .....	49



## DAFTAR LAMPIRAN

Lampiran 1. <i>Flowchart</i> kategorisasi fonem lengkap.....	55
Lampiran 2. Daftar pengucapan fonem.....	55
Lampiran 3. Format kuesioner .....	56
Lampiran 4. Hasil kuisisioner .....	61
Lampiran 5. <i>Database</i> audio.....	63
Lampiran 6. <i>Source code</i> .....	63

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat, hidayah dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir dengan judul "Implementasi Algoritma *Knuth-morris-pratt* Dalam Sistem Pembuatan Korpus bahasa Indonesia dan Integrasi Kategorisasi Fonem Untuk Meningkatkan Akurasi *Text-to-speech*". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan dan nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Allah SWT atas berkat dan Rahmat-Nya sehingga penulis dapat memiliki kehidupan yang baik dan menyelesaikan tugas akhir ini.
2. Kedua Orang tua penulis, Bapak Budi Wijaya, S.Ag., M.Pd. dan Ibu Masirah BSW yang selalu memberikan dukungan, doa dan semangat yang tiada hentinya.
3. Kakak penulis, Fachri Ananda nusantara yang selalu memberikan support dalam perjalanan perkuliahan penulis.
4. Keluarga penulis, Tante Tuti, Om Jaya, Om Herman, Tante Fiha, beserta semua sepupu penulis, telah memberikan tempat dan bantuan kepada penulis dalam beradaptasi di lingkungan baru di Sulawesi Selatan ini.
5. Bapak Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN. Eng. selaku pembimbing I dan Ibu Elly Warni, S.T., M.T. selaku pembimbing II yang senantiasa meluangkan waktu, tenaga, perhatian untuk memberikan saran yang luar biasa dalam mengarahkan penulis sehingga dapat menyelesaikan tugas akhir ini.
6. Segenap staf, dosen dan civitas akademika Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin yang telah membantu kelancaran hingga tugas akhir dapat terselesaikan.
7. Dr. Ikhwan M. Said, M.Hum yang telah memberikan saran dan membuka wawasan terkait fonetik bahasa Indonesia sehingga memberikan langkah awal dalam pengerjaan tugas akhir ini.
8. Kak Dhinda Fitri Wiludjeng yang selalu dengan senang memberikan masukan dalam pengembangan dalam pembuatan sistem korpus ini.
9. A. Rusmiati, Dea Wahsa, Juan Jimmy, Pahrul, Sabda, Dita, Deby, Citra, Besse, Zid yang menjadi tempat bertanya penulis saat penulis mengalami kendala dan masalah terkait tugas akhir.
10. M. Reza, Giga, Agil, Wira, teman teman dan kakak kakak lab AI yang secara supportive selalu menyemangati dalam penyelesaian skripsi ini.
11. Leon, Rayyan, Putri, A. Diatri, Yusuf, Ilal, Fatur, Hedar, Farhan dan teman-teman Kelas C yang berbagi kenangan menyenangkan selama berkuliah di sini.
12. Sayyid, Afwan, Zaki dan Teman-teman S19NIFIER yang telah berbagi kenangan suka dan duka selama berkuliah di Teknik Informatika Universitas Hasanuddin.

13. Arlin, Dylan, Nadin, May, Ima, Ian, Anggi, Gaby, Nopal, Angel, aswin, Havidz, Vira dan teman teman UKM PSM Universitas Hasanuddin semuanya, yang telah berbagi kenangan berharga.
14. Teman-teman KKNT Gel.108 Posko 8 Desa Garessi kec. Tanete rilau kab. Barru (Tifa, Huusnul, Asifah, Dinar, Nanda, Nabila, wahyudi) yang memberikan pengalaman baru kepada penulis saat KKN.
15. Teman-teman dari Komunitas Kucing Unhas yang telah menjadi tempat menyegarkan diri serta memberikan pengalaman kepada penulis untuk peduli dan bersikap ramah terhadap hewan, khususnya kucing liar.
16. Dan kepada semua pihak lain yang dengan sadar atau tidak sadar telah membantu penulis, yang tidak dapat penulis sebutkan satu per satu, terima kasih atas telah membantu penulis dalam menyelesaikan tugas akhir ini.

Akhir kata, penulis menyadari bahwa skripsi ini masih jauh dari sempurna dan masukan serta kritik membangun sangat penulis harapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat dan memberikan kontribusi positif dalam pengembangan ilmu pengetahuan di bidang pembuatan korpus fonem bahasa Indonesia. Aminn

Gowa, 27 Desember 2023

Penulis,  
Arif Putera Wijaya

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam era digital yang terus bergerak maju, pembacaan suara otomatis atau *Text-to-speech* (TTS) menjadi sesuatu yang menarik untuk dikembangkan. *Text-to-speech* diaplikasikan di berbagai hal seperti dalam aksesibilitas bagi individu dengan disabilitas, sistem informasi multimedia dan sebagainya. Namun, meskipun *Text-to-speech* telah berkembang pesat, masih ada beberapa kendala dalam akurasi pembacaan. Salah satu masalah tersebut adalah ketimpangan pembacaan yang disebabkan oleh kurang baiknya model suara dalam memahami bahasa.

Setiap bahasa memiliki karakteristik unik yang membedakannya dengan bahasa lain atau bisa disebut dengan kaidah fonem bahasa. Perbedaan ini dapat dilihat dalam fonologi, morfologi, atau sintaksis. Fonologi adalah bagian dari tata bahasa atau bidang linguistik yang menganalisis bunyi bahasa secara umum (Teunomvira & Gustianingsih, 2022). sedangkan fonem adalah unit bunyi terkecil yang tidak dapat diamati secara jelas atau tidak diartikulasikan. Fonem terdiri atas fonem vokal, konsonan dan semi vokal (Darwin dkk., 2021) yang mana fonem berbeda memiliki tata cara baca yang berbeda.

Penelitian yang dilakukan oleh Dhinda Fitri Wiludjeng pada tahun 2022 dengan judul "Pembuatan Korpus Berbahasa Indonesia menggunakan Metode Diphone *Concatenation* Untuk Sistem *Text-to-speech*". Penelitian tersebut memperoleh hasil pembuatan Sistem *Text-to-speech* yang mengkonversi teks menjadi suara dengan data suara sendiri dan menggunakan korpus bahasa Indonesia dengan metode Diphone *Concatenation* (Wiludjeng, 2022). Namun hasil penelitian ini masih kurang dikarenakan penambahan korpus dilakukan secara manual serta bentuk fonem yang digunakan pada korpus masih belum sesuai kaidah bahasa Indonesia, sehingga hasil suara yang dikeluarkan kurang baik. Selain itu terdapat penelitian yang dilakukan oleh Mohammad Ilham pada tahun 2020 dengan judul "Penerapan Algoritma *Knuth-morris-pratt* Dalam Fitur Pencarian Pengarsipan Dokumen Pada SMA Plus Negeri 17 Palembang". Penelitian tersebut memperoleh hasil bahwa Algoritma *Knuth-morris-pratt* berhasil diterapkan pada pencarian

pengarsipan dokumen di SMA Plus Negeri 17 Palembang, penelitian tersebut mencatat rata-rata performa algoritma *Knuth-morris-pratt* (KMP) pada form pencarian arsip dokumen sebesar 0.0017 detik dengan dokumen sebanyak 100 data dokumen (Ilham & Mirza, 2020).

Penelitian ini akan menggunakan algoritma KMP dan integrasi kategorisasi fonem untuk membaca dan menambah pada *database* korpus bahasa Indonesia. Algoritma KMP akan membantu dalam proses pencarian kata-kata dalam korpus dan memastikan bahwa kata-kata yang ditemukan sesuai dengan yang diinginkan. Integrasi kategori fonem sesuai kaidah fonem bahasa Indonesia akan membantu dalam meningkatkan akurasi *Text-to-speech*. Dengan demikian, penelitian ini diharap dapat meningkatkan akurasi sistem *Text-to-speech* melalui sistem pembuatan korpus bahasa Indonesia dengan mengkategorisasi fonem yang ada. Hal ini diharapkan dapat memperbaiki dan mengurangi ketimpangan pembacaan kata yang terjadi pada sistem *Text-to-speech*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut Masalah dalam penelitian ini dirumuskan menjadi beberapa point berikut:

1. Bagaimana implementasi algoritma *Knuth-morris-pratt* dalam sistem pembuatan korpus bahasa Indonesia?
2. Bagaimana integrasi kategorisasi fonem dalam sistem pembuatan korpus bahasa Indonesia untuk *Text-to-speech*?
3. Bagaimana pengaruh integrasi fonem terkategori dalam pembuatan korpus bahasa Indonesia terhadap akurasi pembacaan *Text-to-speech*?

## 1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah untuk :

1. Mengimplementasikan algoritma *Knuth-morris-pratt* untuk mempercepat proses pembuatan korpus bahasa Indonesia.
2. Menganalisis bagaimana kategorisasi fonem diterapkan dalam sistem pembuatan korpus bahasa Indonesia untuk *Text-to-speech*.

3. Menganalisis pengaruh integrasi fonem terkategoriisasi dalam korpus bahasa Indonesia terhadap akurasi pembacaan *Text-to-speech*.

#### **1.4 Manfaat Penelitian**

Adapun manfaat yang diperoleh adalah memberikan sistem pembuatan korpus bahasa Indonesia yang lebih cepat, serta memberikan informasi dan wawasan terkait kategoriisasi fonem terhadap akurasi *text-to-speech*.

#### **1.5 Ruang Lingkup**

Batasan masalah dalam penelitian ini adalah:

1. Bahasa yang digunakan adalah bahasa Indonesia.
2. Pola struktur kata yang digunakan adalah V, VK, KV, KVK, KKV dan VKK.
3. Tidak ada pembeda dalam pembacaan antara ê dan e.
4. Fonem yang terbentuk dibagi berdasarkan jumlah huruf, yaitu fonem 1 huruf, 2 huruf dan 3 huruf.
5. Input kata baru dalam kategoriisasi fonem minimal 3 huruf.
6. Suara yang dihasilkan belum memperhatikan aspek intonasi dan emosi.

## BAB 2 TINJAUAN PUSTAKA

### 2.1 *Text-to-speech*

Sistem *text-to-speech* juga sering di singkat TTS merupakan suatu sistem yang dapat melakukan konversi *inputan* berupa teks menjadi *output* berupa suara. Sistem ini berkembang sejak abad ke-18 tepatnya pada tahun 1759, Professor Ch. G. Kratzenstein menciptakan rongga resonan (*resonant cavities*) yang *diexcite* (diaktifkan) dengan suatu vibrator sehingga dapat menghasilkan suara dari 5 vokal (huruf hidup) yaitu a, i, u, e dan o (Wundari, 2009). Hingga saat ini, TTS telah menghasilkan *output* suara dari *inputan* teks secara otomatis dengan cara *fonetisasi* yang cara kerjanya adalah dengan menyusun fonem-fonem untuk membentuk ucapan. Seiring dengan perkembangan tersebut, TTS saat ini telah diaplikasikan dalam berbagai bidang diantaranya (Widyawicara, 2021):

1. Pendidikan :

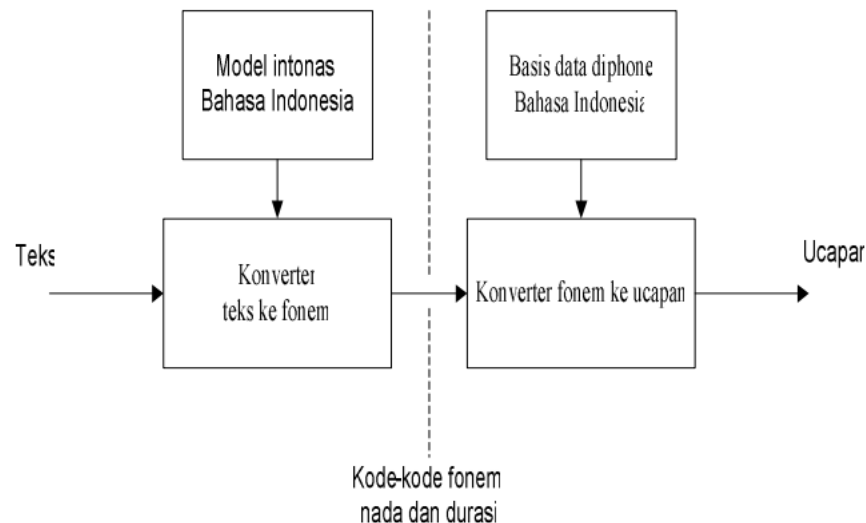
Teknologi TTS memiliki peran yang sangat penting dalam bidang pendidikan. Salah satu manfaatnya adalah membantu siswa dalam mempelajari kata-kata dengan cara mendengarkan pengucapan kata-kata tersebut melalui audiobook. Selain itu, penggunaan audiobook juga dapat meningkatkan kemampuan literasi siswa.

2. Bisnis :

Dalam dunia bisnis, teknologi TTS juga memiliki peranan yang signifikan, terutama dalam membantu perusahaan merespons pertanyaan-pertanyaan umum dari pelanggan melalui *chatbot customer service*. Dengan bantuan TTS ini, pelanggan dapat menerima jawaban yang akurat dan cepat, bahkan di luar jam kerja.

3. Fasilitas Publik :

Teknologi TTS juga dapat dimanfaatkan di fasilitas publik, misalnya saat memberikan pengumuman atau himbauan secara berulang di area publik. Selain itu, penggunaan TTS juga bisa menjadi alat bantu untuk membacakan tulisan-tulisan yang ada di mesin ATM bagi mereka yang memiliki gangguan penglihatan.



Gambar 1. Sistem *text-to-speech*

Sumber: (Hapsari, 2014)

Selain dari apa yang sudah disebutkan sebelumnya, ada banyak penerapan lain dari TTS. Hal ini dikarenakan upaya yang dilakukan dalam pengembangan sistem TTS. Pada umumnya, sistem TTS terdiri dari dua bagian yang dapat dilihat pada Gambar 1 adapun penjelasan kedua bagian tersebut adalah sebagai berikut:

1. Konversi Teks menjadi Fonem :

Bagian ini memiliki peran yang penting dalam mengolah teks yang diberikan dalam bahasa tertentu menjadi serangkaian kode suara yang mewakili bunyi, termasuk durasi dan intonasinya. Kode bunyi ini adalah representasi dari unit-unit suara yang akan diucapkan.

2. Konversi Fonem menjadi Ucapan :

Bagian ini menerima *input* berupa kode-kode bunyi beserta informasi tentang nada dan durasi yang dihasilkan oleh proses konversi sebelumnya. Berdasarkan informasi tersebut, bagian perubahan bunyi menjadi ucapan akan menghasilkan suara atau sinyal ucapan yang sesuai dengan kalimat yang akan diucapkan.

## 2.2 Korpus

Korpus merupakan data bahasa yang telah dikumpulkan, diatur dan disimpan dalam format yang dapat digunakan baik itu tertulis maupun digital dan bertujuan untuk menganalisis linguistik, mengembangkan model bahasa, melakukan



penelitian sastra, ataupun berbagai keperluan lainnya. Secara definisi, Menurut Kamus Besar bahasa Indonesia (KBBI), korpus mengacu pada kumpulan ujaran yang tertulis atau lisan yang digunakan untuk menyokong atau menguji hipotesis tentang struktur bahasa. Demikian pula, menurut beberapa sumber korpus didefinisikan sebagai berikut:

1. Adi Budiwiyanto, seorang Koordinator KKLP Perkamusan dan Peristilahan di Badan Pengembangan dan Pembinaan bahasa menerangkan bahwa korpus adalah kumpulan teks alami, baik bahasa lisan maupun bahasa tulis, yang disusun secara sistematis (Budiwiyanto, 2022).
2. Nur Hizbullah, berpendapat bahwa korpus adalah sekumpulan data baik data biasa ataupun data digital, yang mengandung berbagai jenis informasi seperti kata-kata, struktur, makna dan konteks. Korpus digunakan untuk keperluan penelitian (Hizbullah dkk., 2016).
3. Millatul Islamiyah, berpendapat bahwa korpus adalah kumpulan teks autentik, baik tulis maupun transkrip percakapan dalam jumlah besar yang disimpan secara elektronik (Islamiyah & Fajri, 2019).

Perkembangan teknologi menciptakan pembaharuan informasi dan pengetahuan, hal ini juga membuat berkembangnya data kebahasaan yang awalnya berbentuk manuskrip atau tulisan biasa di atas kertas atau objek lain menjadi seperti sekarang dimana data kebahasaan yang berbentuk teks dan bahasa lisan di ubah menjadi bentuk digital sehingga lebih mudah disimpan dan diolah menjadi berbagai produk seperti kamus ataupun di implementasikan dalam aplikasi pengolah suara. Untuk keperluan itu, format korpus bukan hanya kumpulan data bahasa namun mewakili beberapa bagian bahasa sehingga pembuatan korpus bergantung pada tujuan yang ingin dicapai. Beberapa jenis korpus diantaranya (Wiludjeng, 2022):

1. *General Corpus* atau korpus umum, berisi teks–teks umum yang terdiri bukan satu jenis teks atau subjek. Contohnya adalah *Brown Corpus* terdiri dari 1 juta kata LOB Korpus terdiri dari 1 juta kata.
2. *Specialized Corpus* atau korpus khusus adalah korpus yang berisi teks yang dibuat dengan tujuan yang spesifik, seperti untuk memberikan sebuah topik. Contohnya *Michigan Corpus of Academic Spoken English (MICASE)* (*spoken registers in a US academic setting*) terdiri dari 5 juta kata.

3. *Regional Corpora* atau korpus regional, korpus yang dibangun untuk merepresentasikan keberagaman atau variasi bahasa dari suatu daerah, seperti dialek.
4. *Monolingual Corpus* atau korpus satu bahasa. Seperti namanya korpus ini dibuat hanya terdiri dari satu bahasa saja, biasanya dilengkapi dengan anotasi.
5. *Parallel corpus* atau korpus paralel, korpus berisi kata dari dua korpus *monolingual* yang saling terhubung / berpasangan.
6. *Multilingual corpus* atau korpus multibahasa, sepertihalnya korpus paralel korpus ini terdiri atas beberapa korpus *monolingual* yang saling berpasangan.
7. *Comparable Corpus* atau korpus sebanding adalah korpus pasangan dari dua atau lebih korpus *monolingual* yang mana teksnya berhubungan dengan topik yang sama. Contohnya, *International Corpus of English (ICE)* yang terdiri dari 1 juta kata.
8. *Learner Corpus* atau korpus pelajar adalah korpus teks yang dibuat oleh peneliti bahasa. Contohnya *International Korpus of Learner English (ICLE)* terdiri dari 20 ribu kata.
9. *Diachronic Corpus* adalah Korpus 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. *Linguistik Corpus* adalah korpus yang terdiri atas sejumlah basis data bahasa yang terakumulasi secara sistematis dari berbagai bidang penggunaan bahasa yang mengikuti beberapa metode statistik dan teknik pengambilan data sampel.

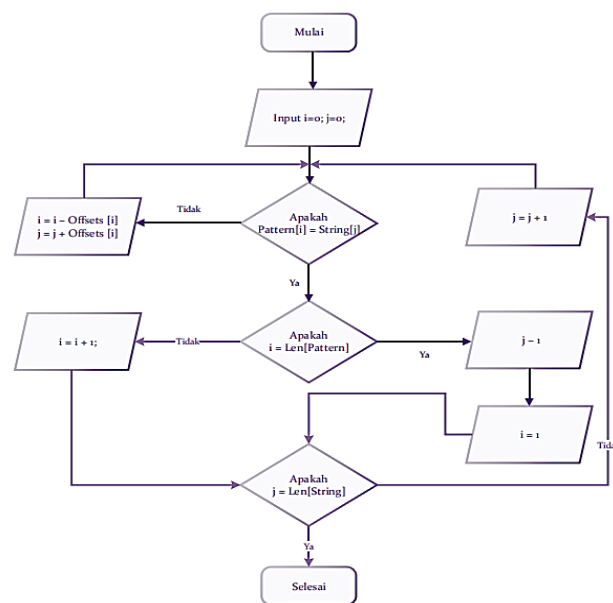
### 2.3 Algoritma *knuth-morris-pratt*

Algoritma *knuth-morris-pratt* atau yang biasa disingkat KMP, merupakan algoritma yang dikembangkan oleh Donald Knuth, Vaughan Pratt dan James H. Morris pada tahun 1977. Algoritma ini merupakan sebuah algoritma pencocokan pola atau *string matching* yang digunakan untuk mencari kemunculan suatu pola (*pattern*) dalam sebuah teks atau *string (text)* (Ilham & Mirza, 2020).

Algoritma ini dikembangkan sebagai solusi terhadap permasalahan pencocokan *string* dengan waktu eksekusi yang lebih efisien dibandingkan metode

pencocokan *string naif* atau *Boyer-Moore*. Metode pencocokan *string Boyer-Moore* akan mencoba semua kemungkinan posisi awal pencocokan pola dalam teks, yang menyebabkan waktu eksekusi yang tidak efisien terutama untuk pola panjang di dalam teks yang panjang. Sedangkan cara kerja algoritma KMP adalah dengan melewati kemungkinan – kemungkinan yang tidak diperlukan untuk menghindari banyaknya jumlah kemungkinan yang ada, dengan demikian waktu yang diperlukan untuk mencari *string* menjadi lebih efisien (Kumara, 2010). *Flowchart* algoritma ini dapat dilihat pada Gambar 2 Berikut adalah saat pencocokan *string*, adapun tahapan yang ada dalam algoritma KMP adalah sebagai berikut:

1. Algoritma KMP melakukan pencocokan *pattern* pada awal teks
2. Pencocokan akan dimulai dari kiri ke kanan teks, algoritma akan mencocokkan karakter per karakter *pattern*, dengan karakter di teks hingga kondisi terpenuhi. Adapun kondisi sebagai berikut :
  - a. Karakter di *pattern* dan teks tidak ada yang cocok (*mismatch*)
  - b. Semua karakter di *pattern* cocok, algoritma akan memberikan posisi teks yang cocok dengan *pattern*.
3. Algoritma menggeser *pattern* berdasarkan Tabel next, lalu menghitung langkah 2 sampai *pattern* berada di ujung teks



Gambar 2. *Flowchart* algoritma *knuth-morris-pratt*

Sumber: (Khairan & Ahmadian, 2019)

## 2.4 Kaidah Struktur Linguistik

Struktur linguistik memiliki bagian-bagian dasar diantaranya adalah fonologi, morfologi, sintaksis, semantik, leksikologi dan sebagainya. Setiap bagian memiliki konsentrasinya masing masing, seperti fonologi yang berkonsentrasi pada persoalan bunyi, morfologi pada persoalan struktur internal kata, sintaksis pada susunan kata dalam sebuah kalimat semantik pada persoalan makna kata dan leksikologi pada persoalan pembendaharaan kata. Adapun kedudukan dalam cabang linguistik sebagai berikut (Muslich, 2018):

1. Fonologi dipelajari dengan melihat 2 sudut pandang yaitu:
  - a. Fonetik, bunyi-bunyi ujar dipandang sebagai media bahasa semata atau bisa dibayangkan sebagai benda atau zat. Hal ini menjadikan bunyi dipandang sebagai bahan mentah seperti batu, pasir, semen, atau bahan lainnya yang digunakan dalam membangun rumah.
  - b. Fonemik, bunyi-bunyi ujar dipandang sebagai bagian dari sistem bahasa. Bunyi digambarkan sebagai unsur-unsur bahasa terkecil yang merupakan bagian dari struktur kata dan sekaligus berfungsi membedakan makna.
2. Morfologi mempelajari struktur internal kata (perilaku kata, proses pembentukan kata, sampai dengan nois yang timbul akibat pembentukan kata), sehingga morfologi diperlukan untuk menjelaskan berbagai kasus penulisan struktur kata sebagai berikut:
  - a. Variasi, kata {pukul} memiliki variasi pembacaan antara [pukUI] dan [pUkUI]
  - b. Sufiks, kata {pukul} menjadi [pukulan] setelah ditambah sufiks {-an}
  - c. Prefiks, kata {pukul} menjadi [memukul] setelah ditambah prefix {meN-}
3. Sintaksis mempelajari mengenai makna tataran kalimat, hal ini dapat diketahui dengan memanfaatkan intonasi, jeda dan tekanan dalam kalimat. Contoh perbedaan makna tataran kalimat adalah sebagai berikut :
  - a. Kalimat perintah, pada kalimat [Kamu di sini!]
  - b. Kalimat tanya, pada kalimat [kamu di sini?]
4. Semantik mempelajari mengenai makna kata, dimana menempatkan kapan kata dapat divariasikan ucapannya dan kapan tidak. Contoh penempatan kata semantik adalah sebagai berikut:

- a. Tidak membedakan makna, seperti kata { duduk } jika dibaca [dudUk] atau [dUdUk].
  - b. Membedakan makna, seperti kata { tahu } jika dibaca [tahU] dan [tAhU].
5. Leksikologi dan leksokografi mempelajari persoalan pembendaharaan kata suatu bahasa baik dalam rangka penyusunan kamus maupun bukan.
  6. Dialektologi mempelajari pemetaan wilayah pemakaian dialek atau variasi bahasa tertentu. Terutama variasi-variasi ucapan pemakaian bahasa, baik secara sosial maupun geografis.
  7. Linguistik terapan mempelajari mengenai pengajaran bahasa kedua dan bahasa asing, yang bertujuan mempelajari keterampilan berbahasa lisan bahasa asing.
  8. Psikolinguistik mempelajari perkembangan penguasaan bunyi, hal ini untuk mengetahui alasan mengapa bunyi bilabial, lateral, dan vocal rendah-depan lebih dahulu dikuasai dibandingkan bunyi labiodental, tril dan vocal tinggi-belakang.

Sejarah perkembangan kajian fonetik dimulai dari terbentuknya *International Phonetic Association* (IPA) pada tahun 1886, meskipun buku-buku yang membahas tentang bunyi sudah ada sejak tahun 1569. IPA mulai melakukan pengkajian fonetik secara intensif pada tahun 1907 setelah Universitas London mengakui usaha daniel Jones, seorang pakar fonetik dari Inggris. Perkembangan fonetik mengalami kemajuan yang signifikan pada tahun 1930-an, terutama dengan dimulainya kongres-kongres yang mengumpulkan para ahli fonetik di seluruh dunia. Beberapa kongres yang penting antara lain:

1. Kongres Ilmu Fonetik Internasional pertama diadakan pada tahun 1932 di Amsterdam.
2. Kongres Ilmu Fonetik Internasional kedua diadakan pada tahun 1935 di London, di mana dihadiri oleh 262 ahli dari 29 negara.
3. Kongres Ilmu Fonetik Internasional ketiga diadakan pada tahun 1938 di Ghent, Belgia, dengan jumlah peserta sebanyak 273 ahli, 18 lembaga dan persatuan, 38 universitas, serta diwakili oleh 32 negara.

Hingga saat ini, terjadi berbagai perkembangan di bidang fonetik, mulai dari studi tentang pengucapan hingga pembentukan berbagai korpus. Sebagai contoh,

menurut J.D. O'Connor, fonetik dapat didefinisikan sebagai ilmu yang terkait dengan suara-suara ujar yang dihasilkan oleh ucapan manusia, yang kemudian dapat diformulasikan menjadi rangkaian simbol yang memiliki fungsi untuk menyampaikan pesan tertentu. Sebagai contoh, kata "buku" dapat direkonstruksi oleh pendengar menjadi rangkaian bunyi [b][u][k][u].

Adapun kaidah dasar struktur bahasa yang dijelaskan dalam buku pedoman ejaan bahasa Indonesia dan buku fonologi bahasa Indonesia diantaranya sebagai berikut :

#### 1. Huruf Vokal dan Huruf Konsonan

Huruf abjad terdiri atas 26 huruf yang terbagi atas 5 huruf vokal dan 21 huruf konsonan. Huruf vokal dan konsonan memiliki bentuk pembacaan seperti yang ditunjukkan pada Tabel 1. dan Tabel 2. (Atmanegara, 2021).

Tabel 1. Huruf vokal

Huruf Vokal	Contoh Pemakaian Dalam Kata		
	Posisi Awal	Posisi Tengah	Posisi Akhir
a	api	padi	lusa
e*	enak	petak	sore
	ember	pendek	-
i	emas	kena	tipe
	itu	simpan	murni
o	oleh	kota	radio
u	ulang	bumi	ibu

Sumber: (Atmanegara, 2021)

Tabel 2. Huruf konsonan

Konsonan	Contoh Pemakaian Dalam Kata		
	Posisi Awal	Posisi Tengah	Posisi Akhir
b	bahasa	sebut	adab
c	cakap	kaca	-
d	dua	ada	abad
f	fakir	kafan	maaf
g	guna	tiga	gudeg

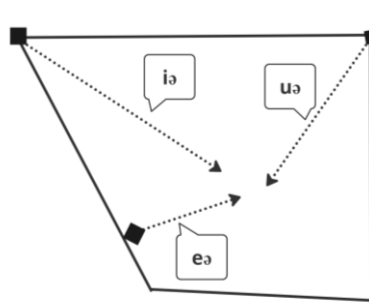
Konsonan	Contoh Pemakaian Dalam Kata		
	Posisi Awal	Posisi Tengah	Posisi Akhir
h	hari	saham	tuah
j	jalan	manja	mikraj
k	kami	paksa	politik
l	lekas	alas	akal
m	maka	kami	diam
n	nama	tanah	daun
p	pasang	apa	siap
q*	qariah	iqra	-
r	raih	bara	putar
s	sampai	asli	tangkas
t	tali	mata	rapat
v	variasi	lava	molotov
w	wanita	hawa	takraw
x*	xenon	-	-
y	yakin	Payung	-
z	zeni	lazim	juz

Sumber: (Atmanegara, 2021)

## 2. Huruf Diftong

Diftong atau perangkapan bunyi vokoid merupakan proses ketika dua deret vokoid diucapkan dalam satu hembusan suara. Ada 2 macam bunyi diftong yaitu (Muslich, 2018) :

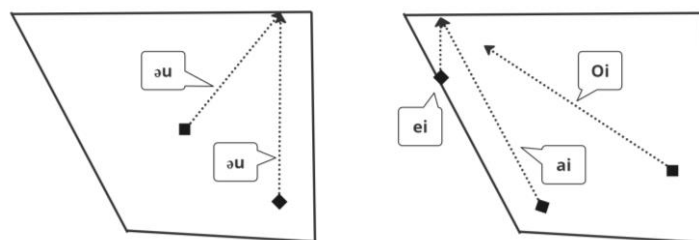
- a. Diftong Menurun (*falling diphthong*), yaitu ketika vokoid pertama bersonoritas sedangkan vokoid kedua kurang bersonoritas bahkan mengarah ke nonvokoid. Kemungkinan terjadinya diftong menurun dapat dilihat pada Gambar 3 (Cth : [a<sub>u</sub>ra] : “aura”).



Gambar 3. Diftong menurun

Sumber: (muslich, 2018)

- b. Diftong menaik (*rising diphthong*), yaitu ketika vokoid pertama kurang bersonoritas atau mengarah ke nonvokoid sedangkan vokoid kedua bersonoritas. Pada Gambar 4 terlihat bahwa kondisi menaik adalah huruf vokal. (Cth : [koI] : “koi”).



Gambar 4. Diftong menaik

Sumber: (muslich, 2018)

### 3. Konsonan Gabungan

Merupakan pengucapan huruf konsonan yang dibaca dalam satu bunyi, dalam bahasa Indonesia ada beberapa gabungan yang sering digunakan dapat dilihat pada Tabel 3. (Atmanegara, 2021).

Tabel 3. Konsonan gabungan

Konsonan Gabungan	Contoh Pemakaian Dalam Kata		
	Posisi Awal	Posisi Tengah	Posisi Akhir
kh	khusus	akhir	tarik
ng	ngarai	bangun	senang
ny	nyata	banyak	-
sy	syarat	musyawarah	arasy

Sumber: (Atmanegara, 2021)



#### 4. Kata Berimbuhan

Merupakan kondisi ketika kata yang mengandung tambahan imbuhan, seperti awalan, sisipan, akhiran dan kombinasi awalan serta akhiran. Banyak dari kata-kata berimbuhan ini terdiri dari tiga huruf yang dibaca secara bersamaan. Hal ini menjadi dasar dalam menetapkan aturan pembacaan fonem tiga huruf dalam program. Beberapa contoh imbuhan yang dapat membentuk fonem tiga huruf dapat dilihat pada Tabel 4 (Atmanegara, 2021):

Tabel 4. Kata berimbuhan

Fonem 3 Huruf																				
ka	ng	ny	sy	ny	ny	ty	sy	te	be	pe	pe	pr	tu	te	pa	ve	su	me	la	ka
n	a	a	a	i	o	a	u	r	r	r	m	i	r	s	n	i	r	n	h	h

Sumber: (Atmanegara, 2021)

#### 5. Kesenyapan (Jeda)

Merupakan kondisi ketika pemutusan suatu arus bunyi-bunyi segmental ketika di ujarkan oleh penutur. Jeda ditandai dengan simbol palang rangkap memanjang (#), sedangkan kesenyapan antara suku kata ditandai dengan palang tunggal (+). Sehingga kalimat jika di transkrip dengan memperhatikan kesenyapan akan terlihat sebagai berikut (Muslich, 2018):

“Ini buku” : [ i+ni#bu+ku ]

Maka dari itu, format transkrip dan pemberian jeda pada space ini dijadikan acuan penulisan korpus serta untuk meningkatkan kealamian *output* yang dihasilkan program.

#### 6. Silaba atau Suku Kata

Satuan bunyi yang terdiri dari satu atau lebih huruf yang mengandung satu inti vokal. Dalam bahasa Indonesia, suku kata terbentuk oleh kombinasi konsonan dan vokal. Setiap kata terdiri dari satu atau lebih suku kata. Terdapat dua teori yang digunakan para linguist atau fonetis dalam mengartikan silaba yaitu :

- Teori Sonoritas, menjelaskan bahwa rangkaian bunyi yang diucapkan penutur selalu memiliki puncak kenyaringan (sonoritas) diantara bunyi yang diucapkan. Misal kata [mendaki] puncak kenyaringannya adalah [e] pada [men], [a] pada [da] dan [i] pada [ki]. Dengan demikian kata [mendaki] memiliki 3 suku kata.

- b. Teori Prominans, menitikberatkan pada gabungan sonoritas dan ciri-ciri suprasegmental, terutama jeda. Ketika rangkaian bunyi diucapkan selain satuan kenyaringan bunyi juga terasa adanya jeda diantaranya, yaitu kesenyapan sebelum dan setelah puncak kenyaringan. Batasan diantara bunyi dilambangkan dengan palang tunggal [+]. Sehingga [mendaki] ditranskrip menjadi [men+da+ki] sehingga memiliki 3 suku kata.

Berdasarkan kedua teori di atas diketahui bahwa sebagian besar suku kata terdiri atas satu bunyi sonor yang berupa vokoid (V) baik tidak didahului dan diikuti kontoid (K), didahului dan diikuti kontoid, didahului kontoid saja, atau diikuti kontoid saja. Pernyataan itu ditulis dalam persamaan (1) berikut :

$$(K) V (K) \quad (1)$$

Adapun struktur suku kata, dalam bahasa Indonesia terdapat 11 pola struktur kata yang terdiri atas vokal (V) dan konsonan (K), seperti yang dijelaskan dalam Tabel 5 berikut (Muslich, 2018):

Tabel 5. Struktur suku kata bahasa indonesia

Struktur Suku Kata	Contoh
V	[a] pada [a+ku] [i] pada [i+bu]
KV	[si] pada [si+ku] [pa] pada [pa+da]
VK	[em] pada [em+ber] [ah] pada [ah+li]
KVK	[tam] pada [tam+pan] [ber] pada [ber+sih]
KKV	[pro] pada [pro+tes]
KKVK	[prak] pada [prak+tis]
KKVKK	[pleks] pada [kom+pleks]
VKK	[eks] pada [eks+por]
KVKK	[seks] pada [seks]
KKKV	[stra] pada [stra+te+gi]
KKKVK	[struk] pada [struk+tur]

Sumber: (Muslich, 2018)

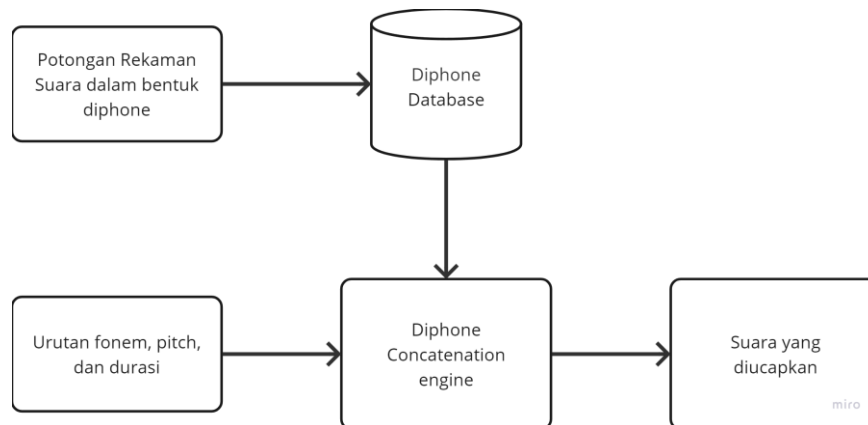
Meskipun terdapat banyak pola namun hanya pola V, VK, KV, KVK yang merupakan pola suku kata asli bahasa Indonesia, sedangkan lainnya merupakan pola suku kata dari bahasa asing dan bahasa daerah yang sudah diadopsi dalam bahasa Indonesia (Suprihatin, 2005).

Mengingat enam dasar aturan di atas, beserta batasan masalah yang ada, yaitu mengenai penggunaan fonem yang terbatas hanya pada fonem yang terdiri dari satu, dua dan tiga huruf, berikut adalah pedoman yang dirancang dalam sistem

1. Apabila ditemukan pola dua vokal berurutan selain diftong, maka memenggalan dilakukan diantara kedua vokal.
2. Apabila ditemukan pola dua konsonan berurutan selain konsonan gabungan, maka dilakukan pemenggalan diantara konsonan tersebut.
3. Apabila ditemukan suku kata konsonan tunggal, maka diberikan tanda “~” yang menandakan pembacaan hanya dibaca ½.
4. Spasi ditandai dengan simbol “#”.
5. Pola yang digunakan adalah pola bahasa Indonesia asli dan imbuhan, yaitu V, VK, KV, KVK, KKV dan VKK.
6. Pola tiga huruf diambil dari kata imbuhan dan juga gabungan antara fonem, contoh : [pergi]:[per+gi], [misal] : [mi+sa+l~], [tampan]:[ta+m~+pan].

## **2.5 Concatenation Synthesizer**

*Concatenation synthesizer* merupakan proses yang mampu menghasilkan sinyal ucapan otomatis menggunakan transkrip grafemke- fonem (Atmaja, 2011), sehingga dengan proses ini memungkinkan menghasilkan *output* suara dari kata atau kalimat dengan cara menggabungkan (*Concatenation*) potongan fonem atau unit suara yang telah direkam sebelumnya, unit suara tersebut disimpan dalam *database* dalam bentuk gelombang (*uncoded*) atau dalam bentuk *encoded* yang disesuaikan dengan pengkodean ucapan (*speed coding method*) yang dipakai.



Gambar 5. Proses *concatenation text-to-speech*

Sumber: (Melangi, 2018)

Gambar 5 di atas mengilustrasikan cara kerja sintesis *concatenation* berdasarkan rangkaian bagian ucapan yang telah direkam. Dasar dari perangkaian sintesis adalah dengan menggabungkan segmen-segmen dari gelombang ucapan alami yang tersimpan dalam *database*. Segmen-segmen tersebut dapat berupa kata-kata (*words*), unit subkata (*subword unit*) seperti fonem (*phonemes*), diphone (*diphones*) dan suku kata (*syllables*). Sintesis ucapan perangkaian secara luas digunakan, bekerja dengan prinsip pemilihan unit (*unit selection*). Beberapa sistem populer sintesis perangkaian yang menggunakan prinsip pemilihan unit antara lain unisyn, clunits dan multisyn. (Melangi, 2018).

*Concatenation synthesizer* memberikan keunggulan berupa *database* yang lebih ringan dan pengolahan *database* korpus yang lebih fleksibel dibandingkan metode lainnya. Namun beberapa permasalahan masih belum terpecahkan dengan tuntas yaitu harapan akan tercapainya *Concatenation Synthesizer* yang memenuhi kualitas terdengar jelas (*intelligibility*) dan kealamian (*naturalness*) (Wiludjeng, 2022).

Oleh karena itu, dalam dunia teknologi audio, khususnya pada pengembangan suara sintesis, suara sintesis secara umum harus memenuhi tiga karakteristik utama, yakni tingkat pemahaman (*intelligibility*) untuk memastikan bahwa teks yang dihasilkan dapat dimengerti oleh pendengar, kelancaran pengucapan (*fluidity*) untuk menjamin lancarnya transisi antara suku kata dan kealamian pengucapan

(*naturalness*) untuk memastikan bahwa teks yang dihasilkan terdengar serupa dengan pengucapan manusia pada umumnya (Melangi, 2018).

## 2.6 Metode Evaluasi *Black box* dan *Mean Opinion Score*

Dalam melakukan evaluasi untuk menguji kinerja sistem, terdapat berbagai metode yang dapat diterapkan. Diantaranya adalah metode *black box* dan *Mean Opinion Score* (MOS). Penjelasan mendalam tentang kedua metode ini diberikan sebagai berikut.

### 1. Evaluasi *Black box*

Metode ini dikenal juga sebagai metode fungsional, adalah sebuah teknik pengujian di mana fokus utama adalah pada *input* dan *output* sistem tanpa mempertimbangkan bagaimana sistem tersebut bekerja di dalamnya. Dalam kata lain, tester/responden tidak memerlukan pengetahuan tentang kode internal, struktur, atau cara kerja dari perangkat lunak yang diuji (Wiludjeng, 2022). Metode ini sangat penting karena memungkinkan pengujian sistem secara holistik, dengan memastikan bahwa fungsionalitas keseluruhan sistem sesuai dengan harapan, tanpa memerlukan pemahaman teknis yang mendalam tentang internal sistem.

### 2. Evaluasi *Mean Opinion Score* (MOS)

Metode ini merupakan metode pengujian yang dilakukan secara subjektif dengan cara melakukan pembagian kuesioner pada penguji. MOS memberikan pemahaman yang lebih mendalam tentang bagaimana suara diinterpretasikan dan diterima oleh pengguna, dengan mempertimbangkan faktor subjektif yang tidak dapat diukur secara langsung melalui metode analisis teknis atau matematis. Adapun perhitungan yang digunakan ditunjukkan pada persamaan (2) berikut (Wiludjeng, 2022):

$$MOS = \frac{\text{jumlah nilai yang diberi responden}}{\text{jumlah responden}} \quad (2)$$