BAB II
LANDASAN TEORI

2.1 Sistem Biometrik

Sebuah sistem biometrik pada hakikatnya merupakan sebuah sistem pengenalan pola yang melakukan identifikasi personal dengan melakukan keotentikan dari karekteristik *fisiologis* (retina, iris, sidik jari, struktur wajah, pola mata, geometri telapak tangan) dan perilaku tertentu yang dimiliki seseorang. Secara logika sistem ini dibagi menjadi dua modul: modul pendaftaran dan modul identifikasi. Modul pendaftaran yang berfungsi untuk mengambil data-data dari individu dan menyimpannya kedalam sistem. Untuk dapat digunakan dalam proses pencocokan, representasi digital tersebut diproses lebih lanjut untuk mendapatkan representasi yang cukup untuk mewakili yang disebut dengan *template. Template* ini kemudian dapat disimpan kedalam *database* didalam sistem biometrik atau pada kartu magnetik *(smart card).* Modul identifikasi berfungsi untuk mengidentifikasi individu pada titik akses. Pada saat pengoperasian, pemindahan biometrik menangkap karekteristik yang akan di identifikasi dan di ubah menjadi format digital, kemudian oleh *extraction feature* diproses menjadi representasi yang sama dengan *template*-nya dan kemudian dicocokan untuk mendapatkan identitas.

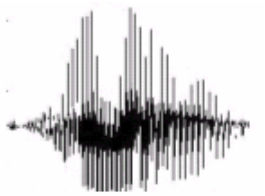
Karekteristik *fisiologis* manusia dapat dipakai sebagai identitas atau identitas personal, dalam karekteristik *fisiologis* manusia untuk biometrik harus memenuhi bebarapa kriteria yaitu: unik, permanen, dan dapat diambil atau diukur

secara kuantitatif dan nilai dari pengukuran tersebut dapat digunakan untuk kepentingan identifikasi.

Beberapa karekteristik *fisilogis* untuk identifikasi personal antara lain yang sudah sering dipakai maupun dalam penyelidikan atau pengembangan lebih lanjut antara lain adalah: wajah, pola retina, *hand geometri*, iris mata, pola suara, tandatangan dan sidik jari.

2.1.1 Pola Suara

Salah satu hal yang pendekatan biometrik sudah banyak dipakai dengan pengenalan suara atau ucapan. Suara dapat di identifikasikan melalui jarak jauh dengan sistem *vocal.* Sistem *vocal* adalah suatu sistem yang digunakan dalam menganalisa suara dengan teknologi *speech recognition.*



Gambar 2.1 Pola suara

2.1.2 Wajah

Wajah dapat diidentifikasi secara otomatis dengan suatu sistem yang kompleks karena tampilan struktur wajah dapat berubah-ubah. Tampilan struktur wajah dapat dilihat dari ekspersi wajah, gaya dan warna rambut, bulu-bulu diwajah telinga, posisi kepala, dan lain-lain yang dapat menghasilkan citra yang berbeda dengan citra yang diambil sebelumnya. Bentuk wajah dapat dilihat pada gambar berikut :



Gambar 2.2 Wajah

2.1.3 Geometri Struktur Tangan

Hand geometry adalah struktur geometri tangan seseorang. Struktur ini termasuk lebar tangan orang, lebar telapak tangan, tebal tipisnya telapak tangan, panjang jari tangan dan lain-lain. Geometri struktur tangan dapat dijadikan sebagai sistem biometrik karena tangan bersifat unik. Bentuk *scane hand geometry* dapat dilihat pada gambar



Gambar 2.3 *Hand Geometry*

2.1.4 Iris Mata

Iris (selaput mata) pada mata dapat dijadikan sebagai sistem biometrik. Struktur yang terdapat pada iris mata ternyata unik untuk setiap manusia, sedangkan pola iris mata sebelah kiri dan kanan berbeda yang letaknya terlindung dibelakang selaput mata (*kornea)* yang membuat iris mata terlindung dari kerusakan. Maka iris mata dapat dilihat pada gambar dibawah ini



Gambar 2.4 Iris Mata

2.1.5 Verifikasi Tandatangan

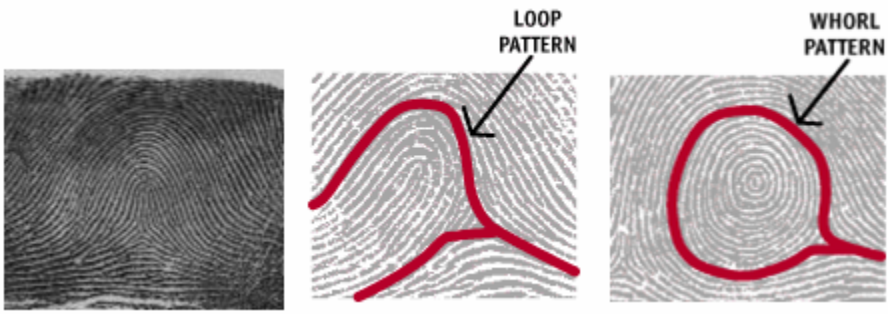
Suatu hal yang mendasari tanda tangan adalah antara bagian yang konstan dengan kebiasaan karena bagian yang berubah-ubah pada tiap kali mentanda tangani. Proses yang mengacu pada aspek dari proses dinamis pembuatan tandatangan, seperti tekanan dan kecepatan tandatangan. Bentuk tandatangan dapat dilihat pada gambar dibawah ini :



Gambar 2.5 Verifikasi Tandatangan

2.1.6 Sidik Jari

Sidik jari dapat dijadikan sebagai sistem biometrik dimana sidik jari bersifat unik. Berdasarkan klasifikasi pola sidik jari dapat dinyatakan secara umum kedalam tiga kategori yaitu: *arch, loop, dan whorl.* Pada pola sidik jari dapat dilihat pada gambar



Arch *Loop* *Whortl*

Gambar 2.6 Pola sidik jari

2.2 Produksi Suara

Suara adalah produk akustik yang spontan dan dikontrol oleh pergerakan sebuah mekanisme vokal manusia. Selama menghasilkan suara, udara dihirup kedalam paru-paru dengan kontraksi tulang rusuk dan membawanya melalui rongga hidung, velum dan trakea. Kemudian dikeluarkan kembali ke udara dengna kontruksi tulang rusuk dan meningkatkan tekanan pada paru-paru. Selama pelepasan udara, udara melewati paru-paru dan melalui senar vokal yang merupakan dua buah kepingan tulang (*ligament*) dan otot-otot yang terdapat di *larynx* dalam trakea. Suara dihasilkan dari getaran senar vokal. Siklus relaksasi diulangi dengan menghasilkan *frekwensi* 80Hz-300Hz. Penghasilan *frekwensi* ini bergantung pada umur seseorang, jenis kelamin, tekanan dan emosi.

Sinyal suara adalah sinyal yang memiliki variasi waktu yang memiliki karakteristik-karakteristik sinyal yang mempresentasikan perbedaan bunyi suara yang dihasilkan. Ada tiga cara untuk memberi label event dalam suara. Pertama adalah bagian diam yang didalamnya tidak ada suara yang dihasilkan. Bagian kedua tidak ada ucapan (*unvoiced state*) dimana didalamnya senar vokal tidak berfibrasi, dengan demikian output dari bentuk gelombnag suara tidak periodik dan benar-benar random. Bagian terakhir adalah bagian yang bersuara, didalamnya memiliki senar vokal yang berfibrasi secara periodic saat udara keluar dari paru-paru.



Gambar 2.7 Bentuk Gelombnag Suara dan Pembagiannya

Suara dihasilkan sebagai bunyi berurutan. Tipe bunyi diproduksi bergantung pada bentuk (*shape*) *vocal tract. Vokal* track berawal dari membukanya senar vokal dan berakhir pada bibir. Area lintasannya bergantung pada posisi lidah, bibir, rahang, langit-langit (*velum*). Oleh karena itu lidah, bibir, rahang dan *velum* memainkan peranan penting dalam produksi suara.

2.3 Pendengaran dan Penjelasaannya

Bunyi yang dapat didengar ditransmisikan ke telinga manusia melalui getaran partikel-partikel dalam udara. Telinga manusia terdiri atas tiga bagian, telinga bagian luar, telinga bagian tengah dan telinga bagian dalam. Fungsi telinga bagian luar untuk mengarahkan varias tekanan udara ke arah gendang telinga dimana bagian tengha telinga merubah variasi tekanan ke dalam gerakan mekanis. Gerakan mekanis ini kemudian ditransmisikan ke telinga bagian dalam, transformasi gerakan-gerakan ke dalam potensial listrik yang lewat melalui urat syaraf pendengaran, korteks dan kemudian ke otak.

2.4 Translator Bahasa Alami ke Bahasa Buatan

Translator bahasa alami ke bahasa buatan yaitu translator yang mengubah perintah -perintah dalam bahasa alami menjadi bahas a buatan yang dapat dieksekusi oleh mesin atau komputer. Sebagai contoh, translator yang memungkinkan kita memberikan perintah bahasa alami kepada komputer. Dengan sistem seperti ini, pengguna sistem dapat memberikan perintah dengan bahasa sehari-hari, mis alnya, untuk menampilkan semua file, pengguna cukup memberikan perintah ”komputer, tolong tampilkan semua file !” Translator akan mentranslasikan perintah bahasa alami tersebut menjadi perintah bahasa formal yang dipahami oleh komputer, yaitu ”dir *.* <ENTER>”.

2.5 Voice Command Control

Salah satu implementasi dari speech recognition adalah *voice command control.* Pada dasarnya adalah sebuah aplikasi yang menggantikan fungsi keyboard dan mouse dengan suara yang diberikan oleh seorang pengguna berupa sebuah kalimat perintah yang ditujukan untuk komputer, dalam menjalankan aplikasi yang ada pada komputernya. Ketika suara dapat dikenali oleh sistem maka sistem Komputer akan mengeksekusi suatu perintah, sesuai perintah yang diberikan oleh penggunanya.

2.6 Pemrosesan Suara

Bentuk gelombang suara harus dikonversi kedalam bentuk digital sebelum diproses dalam sistem pengenalan suara. Bentuk gelombang yang masih asli harus dalam bentuk analog sebelum dikonversi. Konversi sinyal analog ke sinyal digital

meliputi tiga fase yaitu fase sampling, quantisasi dan fase pengkodean. Dalam fase sampling, sinyal analog dikonversikan dari bentuk gelombang kontinyu dalam satuan waktu ke bentuk sinyal diskrit . Sinyal diskrit mengacu pada urutan sample yang memiliki waktu diskrit.

Fase quntisasi, perkiraan nilai sebuah variabel dikonversi ke dalam satu nilai yang terbatas terdapat dalam kunpulan kode . Dalam dua bagian ini menyebabkan bentuk gelombang suara dapat direpresentasikan oleh urutan nilai - nilai memiliki sekumpulan nilai terbatas.

Setelah melewati fase sampling quantisasi sinyal kemudian dikodekan dalam fase pengkodean. Sin yal biasanya direpresentasikan dalam bentuk kode biner.

2.6.1 Sampling

Menurut Nyquist mempersyaratkan minimun sampling rate yang dibutuhkan adalah dua kali frekuensi sinyal.

$$fs = 2 \times fmax$$

dimana :

fs : Frekuensi sampling

fmax : Frekuensi tertinggi sinyal yang disampel

Minimun sampling frekuensi ini dibutuhkan untuk membangun bentuk gelombang dengan frekuensi terbatas dengan tanpa error.

2.6.2 Quantisasi

Sinyal suara biasanya memakai nilai amplitudo yang mendekati nol pada nilai ekstrim puncak yang diijinkan. Sinyal suara dengan bentuk yang tidak seragam distribusi amplitudonya kemungkinan besar mengalami pengquantisasian noise jika ukurannya tidak direduksi pada nilai amplitudonya mendekati nol dan bertambah pada nilai yang besar secara ekstrim.

2.6.3 Pengkodean Suara

Pengkode suara digital dapat diklasifikasikan ke dalam dua buah kategori utama yaitu bentuk yang dapat diperkirakan. Bentuk gelombang coders menggunakan algoritma untuk mengkodekan dan mendekodekan sinyal suara sehingga menghasilkan output sistem dalam bentuk yang dapat diperkirakan.

Vocoder mengkodekan sinyal suara dengan mengumpulkan sekumpulan parameter-parameter yang didigitalkan dan ditransmisikan ke dalam receiver. Kumpulan parameter-parameter digital ini digunakan untuk mengatur nilai-nilai pada parameter dalam fungsi generator dan filter, yang akan menyatukan output sinyal suara. Vocoder tidak memperhitungkan input bentuk gelombang sinyal dan bisa menghasilkan bunyi yang tidak natural.

2.7 Sistem Speech Recognition

Sistem *Speech Recognition* atau Sistem Pengenalan Ucapan adalah sistem yang berfungsi untuk mengubah bahasa lisan menjadi bahasa tulisan. Masukan sistem adalah ucapan manusia, selanjutnya sistem akan mengidentifikasikan kata

atau kalimat yang diucapkan dan menghasilkan teks yang sesuai dengan apa yang diucapkan.

Pada gambar diatas dperlihatkan konfigurasi suatu sistem pengenalan ucapan. Sinyal ucapan (s(n)) pertama kali akan dilewatkan pada bagian penganalisis ucapan untuk mendapatkan besaran-besaran atau ciri-ciri yang mudah diolah pada tahap berikutnya. Untuk setiap ucapan yang berbeda akan dihasilkan pola ciri yang berbeda.

Sistem *Speech Recognition* biasanya dapat dioperasikan dua mode yang berbeda. Pertama adalah mode belajar. Pada mode ini, sistem akan dilatih dengan menggunakan sejumlah kata atau kalimat yang memenuhi suatu kriteria tertentu.

Kedua adalah mode produksi atau pengenala ucapan. Pada mode ini, setiap kalimat yang ingin dikenali akan dianalisis polanya. Pada prinsipnya teknik-teknik atau algoritma yang digunakan pada sistem pengenal ucapan tidak bersifat sensitif terhadap bahasa. Artinya, sistem yang sama dapat digunakan untuk bahasa apapun.

Sistem yang lebih sederhana adalah sistem yang hanya dapat mengenal sejumlah kata yang jumlahnya terbatas. Sistem ini biasanya lebih akurat dan lebih mudah dilatih, tetapi tidak dapat mengenal kata yang berada di luar kosa kata yang pernah diajarkan.

2.7.1 Sistem Pengenalan Kata

Pada dasarnya cara kerja perangkat keras pada pengenalan kat ini adalah input suara manusia yang masuk akan diterima oleh *mikrofon* dengan tujuan untuk

mengubah sinyal suara menjad sinyal listrik. Karena sinyal *output* dari *mikrofon* sangat lemah, maka diperlukan sebuah rangkaian yang berfungsi untuk memperkuat sinyal dari mikrofon tersebut.

Sinyal analog dari mikrofon disampling dengan frekuensi 8000 Hz. Pemilihanfrekuensi sampling sebesar 8000 Hz karena frekuensi terbesar dari sinyal akan disampling adalah 3400 Hz sesuai dengan frekuensi suara manusia sekitar 600 – 3400 Hz. Hal ini dikarenakan adanya syarat bahwa frekuensi sampling minimal dua kali dari frekuensi sinyal yang aka disampling lama waktu sampling data adalah setengah detik dengan demikian dalam satu kali sampling akan diperoleh 4000 data. Kemudian data ini akan diproses dengan analisa LPC sehingga menghasilakn koefisien-koefisien cepstral yang menjadi parameter untuk dikenali.

2.7.2 Beberapa Pendekatan dalam Pengenalan Suara

Manusia adalah “mesin” yang paling baik untuk mengenali dan memahami suara. Dengan mengkombinasikan variasi yang luas tentang pengetahuan *linguistik* yang meliputi sintaksis semantic dan mengadopsi pengetahuan ini menurut tingkat kesulitan. Sistem pengenalan suara dibuat dengan maksud untuk menandingi atau melebihi kemampuan manusia.

2.8 Markov Models

Misalkan sebuah system dideskripsikan pada setiap waktu sebagai salah satu dari N status {1,2,...,N}. Sistem mungkin mengalami perubahan status dalam

satuan waktu diskrit (mungkin kembali ke status semula) tergantung kepada aturan probabilitas setiap status. Misalkan waktu disimbolkan sebagai t, status pada setiap waktu disimbolkan sebagai qt. Deskripsi keseluruhan probabilitas dari sistem secara umum membutuhkan spesifikasi status saat ini dan status-status sebelumnya.

Untuk kasus khusus dengan spesifikasi waktu diskrit, *first order,* dan *Markov chain,* dapat rumus ketergantungan probabilitas suatu status terhadap status-status sebelumnya disederhanakan hanya tergantung pada tepat satu status sebelumnya, yaitu,

$P[q_t = j \mid q_{t-1} = i, q_{t-2} = k,...] = P[q_t = j \mid q_{t-1} = I]$

Penyederhanaan lebih lanjut dengan hanya memperhitungkan proses di bagian kanan, bebas terhadap waktu, maka pendekatan aturan probabilitas transisi status

$a_{ij} = P[q_t = j \mid q_{t-1} = I]$

dengan sifat

$a_{ij} \geq 0$, untuk setiap j, i

$a_{ij} = 1$, untuk setiap i, i =1 sampai N

Karena mematuhi batasan *stochastic* standar.

Contoh permasalahan :

Cuaca dalam satu hari dimodelkan ke dalam tiga status: rainy (1), cloudy (2) dan sunny (3). Aturan probabilitas dari setiap transisi status dideskripsikan sebagai berikut

$$A = \{ a_{ij} \} = \begin{matrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{matrix}$$

Probailitas cuaca untuk 8 hari berturutturut “sun-sun-sun-rain-rain-sun-cludysun” adalah :

Misalkan O adalah tahap pengamatan O = {sunny, sunny, sunny, rainy, rainy, sunny, cloudy, sunny }= {3,3,3,1,1,3,2,3}

$$P(O \mid Model) = P[3,3,3,1,1,3,2,3 \mid Model]$$

$$= P[3]\, P[3|3]^2\, P[1|3]\, P[1|1]$$

$$= II_3.(a_{33})^2\, a_{31}\, a_{11}\, a_{13}\, a_{32}\, a_{23}$$

$$= (1.0)\ (0.8)2\ (0.1)\ (0.4)\ (0.3)\ (0.1)\ (0.2)$$

$$= 1.536 \times 10^{-4}$$

Dengan,

$$II_3 = p[q_t = I\ ],\ 1< i < N$$

2.8.1 Hidden Markov Models

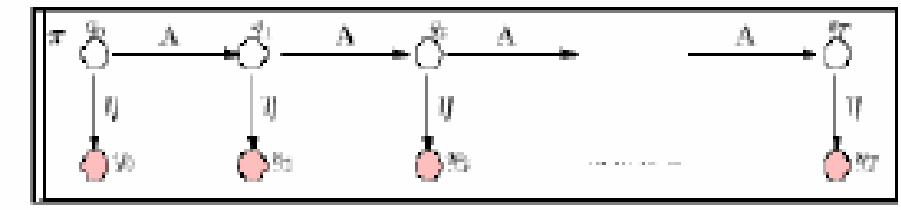
Konsep Markov Model diperluas menjadi Hidden Markov Model dimana observasi merupakan fungsi probabilitas dari status. Model ini dapat diaplikasikan pada kasus yong proses tidak dapat diobservasi secara langsung (tersembunyi) tetapi bisa diobservasi hanya melalui kumpulan proses stokastik yang lain yang menghasilkan tahapan observasi, misalnya persoalan *cointtoss* dan *speech recognition* (pengenalan suara).

HMM merupakan perluasan dari teori yang lebih sederhana yaitu *markov chain* . Markov chain merupakan teori yang menjabarkan kemungkinan terjadinya suatu kejadian berdasarkan rangkaian kejadian sebelumnya dari data data

probabilitas kemunculan setiap kejadian yang mungkin serta probabilitas kemunculan dua kejadian berurutan.

2.8.2 Model Matematika *Hidden Markov Model*

Untuk proses yang dimodelkan dengan HMM status sistem yang sesungguhnya tidak diketahui dan direpresentasikan dengan variabel acak *Hidden Variabel* yang diketahui adalah yang bergantung pada status variabel output. Masalah utama dalam HMM adalah inferensi dimana variabel status yang tidak diketahui diinferensi dari sekuens observasi dan dapat dipecahkan dengan algoritma analisis sinyal. Masalah lainnya adalah *parameter estimation problem*, dimana parameter distribusi kondisional diestimasi dari sekuens observasi. Masalah ini diselesaikan dengan algoritma analisis sinyal. Contoh penggunaan HMM ditunjukkan pada gambar berikut :

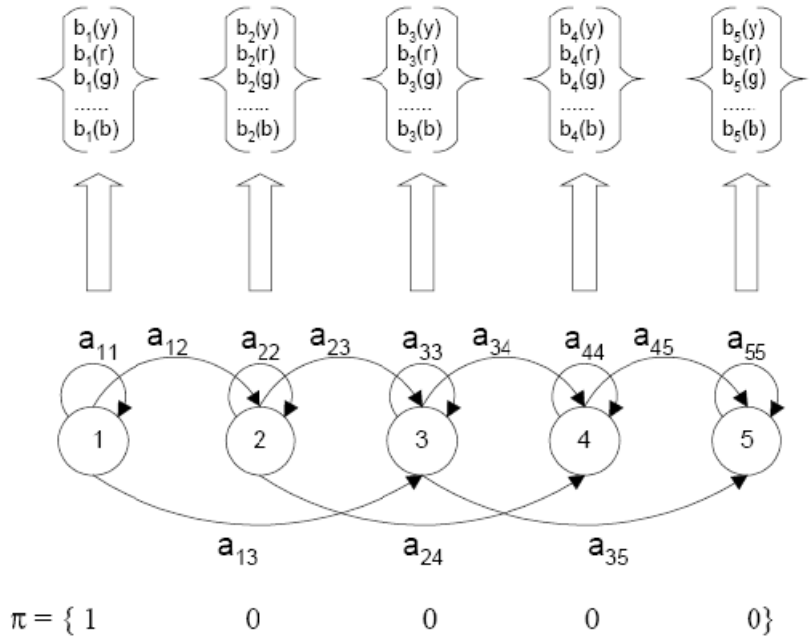


Gambar 2.8 Model Grafik Statistik

HMM direpresentasikan dalam sebuah model grafis statistik. Lingkaran merepresentasikan variabel random. Lingkaran berwarna (*yi*) adalah observasi sedangkan yang tidak berwarna (*qi*) adalah variabel status yang tidak diketahui dan akan diinferensi. *Qi* adalah nama tombol ke-i dalam sekuens dan *yi* adalah kelas penekanan tombol hasil pengelompokkan. Panah dari *qi* ke *qi*+1 dan dari

*qi*ke *yi* mengindikasikan bahwa yang berikutnya bergantung pada kondisi sebelumnya. Nilai pada panah adalah *entry* dari matriks probabilitas A dengan persamaan *p*(*qi*+1|*qi*) = *Aqi*, *qi*+1 yang menunjukkan bahwa tombol *qi*+1 muncul setelah tombol *qi*. Matriks A adalah sebuah cara merepresentasikan data distribusi bigram plainteks dan ditentukan oleh tata bahasa dan diperoleh dari sekumpulan teks bahasa. Terdapat pula persamaan *p*(*yi*|*qi*) = *çqi,yi* , yang menunjukkan probabilitas tombol *qi* dikelompokkan ke dalam kelas *yi* pada langkah sebelumnya. Dengan nilai *yi* yang diketahui dan output matriks *ç* tidak diketahui, kita perlu menginferensi nilai *qi*. Algoritma EM dan Viterbi digunakan untuk mengestimasi parameter (menghasilkan matriks *ç*) dan menginferensi *qi*. Tombol *space* mudah dibedakan oleh pendengaran karena memiliki suara yang unik dan cukup sering digunakan. Penandaan sejumlah tombol *space*, pencarian kelas yang telah dikelompokkan untuk masing-masing tombol, penghitungan estimasi probabilitas untuk setiap anggota kelas dan penyimpanan nilai sebagai *ç* kemudian dilakukan untuk memberi hasil yang baik.

HMM itu adalah suatu teknik perkiraan pola yang probabilistic di mana pengamatan-pengamatan dianggap sebagai keluaran dari proses stokastik dan terdiri dari satu mendasari Rantai Markov. Itu mempunyai dua komponen suatu status(negara yang terbatas Rantai Markov dan suatu yang terbatas



Gambar 2.9 Rantai 5 Data

menunjukkan kemungkinan transisi status(negara (kemungkinan sedang berada dalam status(negara Sj status(negara yang diberi Si)

a_{ij} =P(qt+1=Sj / qt=Si)

bj(wk) adalah simbol wk (warna peluru/bola) sebaran kementakan di dalam suatu keadaan Sj w adalah abjad dan kali 1000 adalah banyaknya lambang di dalam abjad ini.

Alasan untuk menggunakan topologi L-R dari HMM adalah karena struktur nya yang tidak bisa dipisahkan yang dapat model arus sementara dari isyarat ujaran dari waktu ke waktu. Itu bisa tidak sangat jelas nyata bagaimana HMM berhubungan dengan isyarat ucapan modell. Ini bisa dipertimbangkan dengan memperhatikan mekanisme pembentukan bunyi. Cara bicara adalah yang dihasilkan oleh melambat bergeraknya organ bagian badan yang artikulasi.

Pengambilan artikulator-artikulator suara atas suatu urutan dari posisi-posisi yang berbeda dan sebagai konsekwensi menghasilkan arus dari bunyi-bunyi itu bentuk isyarat ucapan. Masing-masing posisi yang artikulasi bisa diwakili oleh suatu keadaan dari yang berbeda dan bermacam-macam jangka waktu. Secara setimpal, transisi antara artikulasi yang berbeda.

2.8.3 Prinsip Kasus dari HMM

Ada tiga kasus utama untuk menangani untuk merumuskan suatu HMM yang sukses. Ini adalah:

Kasus 1 : Evaluasi

Dengan :

- ✓ a suatu model ? = (A, B, p) siap untuk digunakan.
- ✓ Urutan pengamatan O = O_1, O_2, O_3,.....O_{t-1}, O_T
- ✓ Hitung P(O/?) : kemungkinan urutan pengamatan memberi model.

Kasus 2 Pengkodean

Dengan :

- ✓ a suatu model ? = (A, B, p) siap untuk digunakan.
- ✓ Urutan pengamatan O = O_1, O_2, O_3,.....O_{t-1}, O_T

Aksi :

- ✓ Mencari urutan status Q = q_1, q_2, q_3,.....q_{T-1}, q_T yang pasti yang hampir bisa dipastikan itu menghasilkan pengamatan-pengamatan yang menggunakan model yang diberi.

Kasus 3 Pelatihan

Dengan :

- ✓ a suatu model ? = (A, B, p) siap untuk digunakan.
- ✓ Pengamatan latiham O_1^k, O_2^k, O_3^k,.....O_{T-1}^k, O_T^k k di mana kali 1000 adalah banyaknya contoh-contoh untuk pelatihan, model.

Aksi : Set parameter-parameter model untuk memaksimalkan P(O/ ?).

kasus 1 adalah satu prosedur evaluasi sebagai kita sedang mencari-cari untuk menemukan kemungkinan menghasilkan pengamatan yang diberi O oleh suatu model yang diberi l.Ini bisa digunakan untuk menemukan model terbaik di antara banyak yang menghasilkan pengamatan yang diberi.

kasus 2 adalah suatu prosedur pengawasandian untuk mendeteksi atau tidak bersembunyi, menyembunyikan urutan status(negara dari suatu yang diberi pengamatan. Pengamatan-pengamatan itu bisa contoh-contoh pelatihan jika kita ingin belajar perilaku dari tiap status(negara dari aspek yang berbeda, seperti jangka waktu atau spektral karakteristik-karakteristik dari tiap status(negara. Beberapa teknik menggunakan jangka waktu status(negara di dalam evaluasi mereka prosedur dan dalam hal ini pengamatan itu akan contoh ujian untuk mendeteksi negara itu jangka waktu.

kasus 3 adalah prosedur latihan itu kepada pengoptimalan parameter-parameter model untuk memperoleh yang terbaik model bahwa mewakili menunjukkan set tertentu pengamatan-pengamatan kepunyaan satu kesatuan.

2.8.4 Pengenalan Suara Dengan Hidden Markov Models

Sistem pengenalan suara modern secara umum berdasarkan kepada Hidden Markov Models (HMM). Hidden Markov Models merupakan model statistik dimana mempunyai keluaran rangkaian symbol dan kuantitas. Dengan memiliki sebuah model yang memberikan kemungkinan dari rangkaian akustik data yang telah diobservasi dari sebuah atau banyak kata (rangkaian kata) akan dapat menyebabkan sistem bekerja dengan rangkaian kata tersebut sesuai dengan aplikasi aturan.

$$\Pr(\text{word} \mid \text{acoustic}) = \frac{\Pr(acoustic \mid word)\Pr(word)}{\Pr(acoustic)}$$

Dari rangkaian data akustik yang ada, Pr (acoustics) adalah konstan dan tidak dapat diabaikan. Pr (word) adalah merupakan kemungkinan terbesar dari suatu kata. Pr (acoustics | word) masa yang paling terlibat di dalam persamaan dan diperoleh dari Hidden Markov Models.

Untuk dapat digunakan sebagai aplikasi pada dunia nyata, tiga masalah yang mendasar harus dapat diselesaikan.

Diberikan:

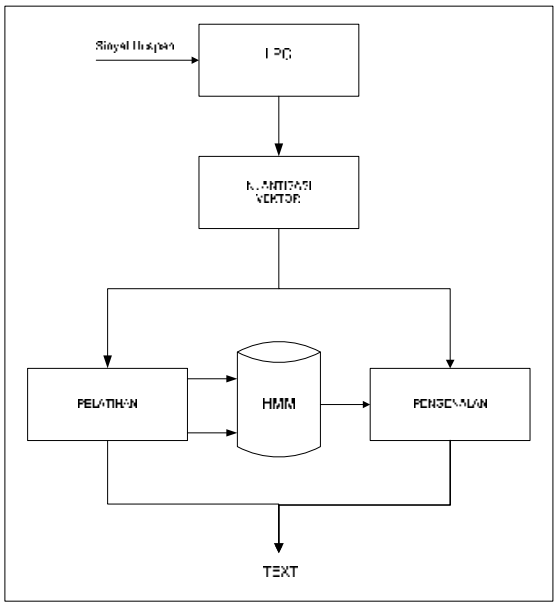
Rangkaian pengamatan:

$O = (o_1\ o_2 \ldots\ldots o_r)$,

Model : $\lambda = (A, B, II)$,

2.8.5 Arsitektur Metode HMM

Arsitektur sistem digambarkan secara global, komponen-komponen yang membangun sistem dan fungsi atau peranannya dalam sistem, serta hubungan antar komponen-komponen tersebut.



Gambar 2.10 Arsitektur *HMM*

Penjelasan Gambar diatas akan dijelaskan sebagai berikut :

1. *Linear Predictive* (LPC) berfungsi untuk mengekstrak sinyal dari data sinyal yang inputkan.
2. Kuantisasi Vektor berfungsi untuk proses pengkodean ciri-ciri vektor sinyal kedalam bentuk simbol HMM. Pengkodean suara digital dapat diklasifikasikan kedalam 2 buah kategori, yaitu bentuk gelombang *coders* dan *vocoders*. Gelombang coders menggunakan algoritma untuk mengkodekan sinyal suara sehingga menghasilkan output sistem dalam bentuk yang dapat diperkirakan. Vocoders mengkodekan sinyal

suara dengan mengumpulkan parameter-parameter yang digitalkan dan ditransmisikan ke dalam receiver.

3. Pelatihan berfungsi untuk pengestimasian ulang parameter HMM dari suatu data berdasarkan data latih yang diajukan pengguna, sehingga hasil estimasi memiliki kualitas yang baik.
4. Pengenal berfungsi melakukan perhitungan probabilitas kemiripan pola dari tiap-tiap model HMM yang mempunyai probabilitas kemiripan yang tertinggi.

2.8.6 *Random Field*

Diberikan $S=\{1,2,\ldots,m\}$ yang menunjukkan *pixel* ataupun *site*. Asumsikan *Ls* himpunan terbatas dari diskrit label untuk tiap $s \in S$ diberikan maka ruang konfigurasi (*configuration space*) χ, dapat didefiniskan untuk setiap $x \in \chi$ sebagai

$$\chi = \prod_{x \in \chi} L_t$$

dimana *x* disebut sebagai konfigurasi.

Definisi 1. Sebuah ukuran probabilitas (*probability measure*) atau distribusi pada χ didefinisikan sebagai sebuah vektor v = (v(*x*)) serta memiliki dua sifat, yaitu:

(i) $v(x) \geq 0 \ \forall \ x \in \chi$,

(ii) $\sum_{x \in \chi} v(x) = 1$

Sebuah *random field* didefinisikan sebagai sebuah distribusi Π pada ??dengan sifat positif mutlak (*strictly positive*), yaitu

$\Pi(x) > 0, \forall x \in \chi,$

2.8.7 *Markov Random Field*

Definisi 2. Misalkan N menunjukkan koleksi himpunan bagian dari *S* dengan relasi

sebagai berikut

$$N := \{N(s) \subset S \mid s \in S\}$$

N disebut daerah sekitar (*neighbourhood*) dari sistem *S* jika memenuhi

(i) $s \notin N(s), \forall s \in S$

(ii) $\forall s, \in t\ S$ dengan $s \in t$ maka $s \in \Leftrightarrow N(t) t \in N(s)$

Site $t \in N(s)$ disebut tetangga (*neigbour*) dari *s*, dinotasikan sebagai <*s,t*> jika *s* dan *t* saling bertetangga.

Pada kasus regular *site grid* , himpunan *S* dapat dianggap sebagai himpunan bagian dari *ZxZ*, yaitu $S=\{(i,j) \in ZxZ \mid -m£i,j£m\}$ untuk lattice bujur sangkar dengan ukuran *(2m+1)x(2m+1)*, sedangkan daerah sekitar dari sebuah *site (i,j)* pada sebuah grid merupakan site-site di sekitar site yang dianggap sebagai titik tengah dengan radius *c*, yaitu

$$N(i,j) := \{(k,l)?\in S \mid 0<(k-i)^2+(l-j)^2 \Leftrightarrow c^2\}.$$

Namun demikian, secara umum radius lingkaran *c* biasanya tidak digunakan untuk menunjukkan order dari daerah sekitar, melainkan digunakan order yang meningkat sejalan dengan nilai

2.8.8 Sifat *Markov* (*Markovianity*) pada *Random Fields*

Markov Random Field merupakan generalisasi dari proses Markov yang terdapat pada teori proses stokastik. Perbedaannya terletak pada domain ketergantungan dari *random variable* yang akan didefinisikan, dimana proses Markov hanya memperlihatkan ketergantungan terhadap waktu pada satu dimensi saja, sedangkan *Markov Random Field* menunjukkan ketergantunan secara spasial. *Definisi 4.* Diberikan sebuah daerah sekitar sebarang N. *Random field P* pada χ disebut *Markov Random Field (MRF)* terhadap sistem daerah sekitar N jika untuk seluruh *site sÎS* karakteristik lokalnya memenuhi persamaan di bawah ini :

$P(Xs\ xs \mid Xr=xr,\ r \neq s) = P(\ Xs=xs \mid Xr=xr,\ r \in N(s))$

yaitu, karakteristik lokal dari sebuah *site* tunggal hanya tergantung pada nilai-nilai (label) dari daerah sekitarnya saja. Besag membuktikan pada (Besag, 74) bahwa sebarang *random field P* ditentukan secara unik oleh karakteristik lokal dari *site* tunggalnya, persamaan (1). Teori ini indah tetapi secara praktis sangat sulit untuk menggunakan *random field* secara umum, karena metode untuk mendapakan distribusi gabungan dari nilai karakteristik lokalnya tidak diketahui. Lebih lanjut, bila diberikan sebuah fungsi, sangatlah sulit untuk menunjukkan apakah fungsi ini merupakan karakteristik lokal dai sebuah *random field* .

2.9 Metode Ekstraksi Sinyal dengan LPC (Linear Predictive Coding)

Metode yang digunakan dalam sistem ini adalah metode *Linear Predictive Coding* (LPC). LPC digunakan untuk mengekstrak pola kata dari suara yang

diinputkan. Ide dasar dari LPC bahwa sebuah sinyal suara manusia pada waktu *n*, *s(n),* dapat diperkirakan sebagai kombinasi linier dari *p* sinyal suara manusia sebelumnya. Hal ini dipresentasikan dengan menggunakan persamaan berikut :

$$S(n) \tilde{} \; a_1.s(n-1) + a_2.s(n-2) + ... + ap.s(n-p) \qquad (1)$$

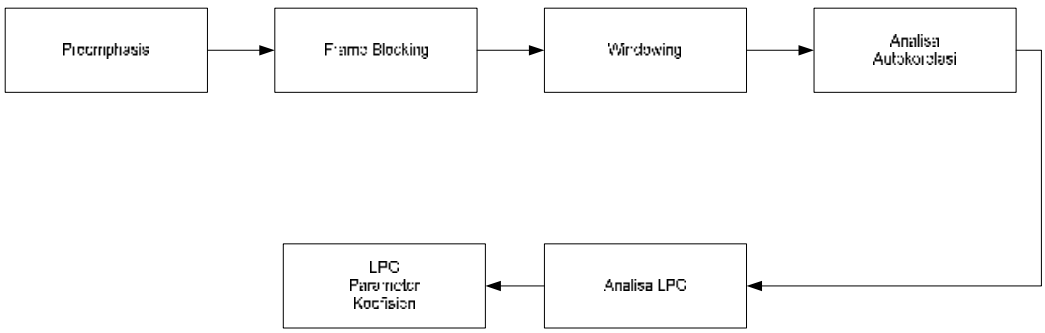
Dimana koefisien-koefisien a1, a2, …., a*p* merupakan asumsi konstanta dari sebuah bagian dari analisa suara manusia.

Persamaan 1 dapat diubah dengan menambahkan sebuah eksitasi G.*u(n)* sehingga menjadi :

$$s(n) = \sum_{f=1}^{p} a_1 s(n-1) + G.u(n) \qquad (2)$$

dimana *u(n)* adalah eksitasi yang dinormalisasi dan G adalah penguatan dari eksitasi tersebut. Jika persamaan diatas ini diekspresikan dalam domain z maka persamaan tersebut berubah menjadi :

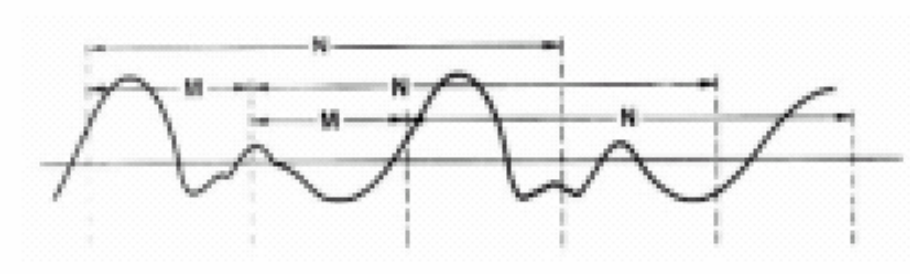
$$S(z) = \sum_{f=1}^{p} a_1.z^{-1}.S(z) = G.U(z) \qquad (3)$$



Gambar 2.11 Diagram Blok LPC

Gambar diatas menunjukkan blok diagram LPC yang biasa digunakan dalam sistem pengenalan suara manusia. Langkah-langkah dasar yang harus dilakukan mengenai blok diagram tersebut adalah :

1. *Preemphasis* : Proses dimana sinyal / speech ucapan dirubah menjadi sinyal.
2. *Blocking into frames* : Pada tahap ini sinyal yang telah di-preemphasis, diblok menjadi beberapa bagian dengan jumlah sample *N*, dan tiap bagian dipisahkan dengan sejumlah M sample
3. *FrameWindowing* : Tahap berikutnya adalah melakukan proses *window* pada setiap bagian sinyal yang telah dibuat sebelumnya. Hal ini dilakukan untuk meminimalkan pada bagian awal dan akhir sinyal. Jika didefinisikan sebuah *window w(n)* dan sinyal tiap bagian adalah x(*n*) maka sinyal hasil proses *windowning*.



Gambar 2.12 *Frame Blocking* sinyal

4. *Auto Correlation Analisys* : Tiap bagian yang telah diberi *window* kemudian akan dibentuk autokorelasinya.
5. Analisa LPC : Langkah berikutnya adalah analisa LPC dimana semua nilai autokorelasinya yang telah dihitung pada tahap sebelumnya akan diubah menjadi parameter LPC
6. Pengubahan parameter LPC menjadi koefisien *cepstral :* Parameter LPC yang sangat penting yang bisa diturunkan dari koefisien LPC adalah koefisien *cepstral* LPC, *c(m).*

2.10 Komputer

Istilah komputer mempunyai arti yang luas dan berbeda bagi setiap orang. Istilah komputer diambil dari bahasa Latin *computare* yang berarti menghitung (*to compute atau to reckon*). Menurut *Blissmer* (1985), komputer adalah suatu alat elektronik yang mampu melakukan beberapa tugas, yaitu menerima input, memproses input sesuai dengan instruksi yang diberikan, menyimpan perintah-perintah dan hasil pengolahannya, serta menyediakan output dalam bentuk informasi.

Sedangkan menurut *Sanders* (1985), komputer adalah sistem elektronik untuk memanipulasi data yang cepat dan tepat serta dirancang dan diorganisasikan supaya secara otomatis menerima dan menyimpan data input, memprosesnya, dan menghasilkan output berdasarkan instruksi-instruksi yang telah tersimpan di dalam memori. Dan masih banyak lagi ahli yang mencoba mendefinisikan secara berbeda tentang komputer. Namun, pada intinya dapat disimpulkan bahwa komputer adalah suatu peralatan elektronik yang dapat menerima *input*, mengolah *input*, memberikan informasi, menggunakan suatu program yang tersimpan di memori komputer, dapat menyimpan program dan hasil pengolahan, serta bekerja secara otomatis.

2.10.1 Central Processing Unit / Processor

CPU merupakan tempat pemroses instruksi-instruksi program, yang pada komputer mikro disebut dengan micro-processor (pemroses mikro). Pemroses ini berupa *chip* yang terdiri dari ribuan hingga jutaan IC. Dalam dunia dagang,

pemroses ini diberi nama sesuai dengan keinginan pembuatnya dan umumnya ditambah dengan nomor seri, misalnya dikenal pemroses Intel 80486 DX2-400 (buatan Intel dengan seri 80486 DX2-400 yang dikenal dengan komputer 486 DX2), Intel Pentium 100 (dikenal dengan komputer Pentium I), Intel Pentium II-350, Intel Pentium III-450, Intel Celeron 333, AMD K-II, dan sebagainya. Masing-masing produk ini mempunyai kelebihan dan kekurangan masing-masing. CPU terdiri dari dua bagian utama yaitu unit kendali (*control unit*) dan unit aritmatika dan logika (ALU). Disamping itu, CPU mempunyai beberapa alat penyimpan yang berukuran kecil yang disebut dengan *register*.

2.10.2 Input/Output Unit

Input/Output Unit merupakan bagian dari komputer untuk menerima data maupun mengeluarkan/menampilkan data setelah diproses oleh *Processor*. Untuk mempermudah pembahasan tentang *Input/Output unit*, pada buku ini akan dijelaskan dalam dua bagian, yaitu: Port I/O dan Peripheral I/O.

1. Port I/O

Port I/O merupakan *Port* atau Gerbang atau tempat dipasangnya konektor dari peralatan I/O, dimana setiap *port* I/O dibawah kontrol dari *Processor*.

a. Port Paralel (LPT1 atau LPT2)

   Merupakan *port* bagi peralatan yang bekerja dengan transmisi data secara paralel. Contoh peralatan yang menggunakan port ini adalah : *Printer*, *Scanner* dll.

b. Port Serial (Com1, Com2)

Merupakan port bagi peralatan yang bekerja dengan transmisi data secara serial. Contoh peralatan yang menggunakan port ini adalah : *Mouse*, Modem , dll.

c. Port AT / PS2

Port ini umumnya digunakan untuk masukan dari *Keyboard, Mouse*.

d. USB Port

USB Port (*Universal Serial Bus*) Port merupakan *Port Serial universal* bagi peralatan yang bekerja dengan transmisi data secara serial. Contoh Perlatan yang menggunakan USB port yaitu Camera Digital.

e. Port VGA

Merupakan port yang berhubungan langsung dengan monitor. Port VGA didapatkan dari pemasangan VGA *Card*.

f. Port Audio

Merupakan port yang berhubungan langsung dengan peralatan audio seperti *Tape, Radio, Speaker*, *Microphone*, dll.

2.10.3 Peripheral I/O

Peripheral adalah sesuatu yang mengacu ke peralatan external yang dihubungkan dengan komputer. Peripheral komputer dapat dibagi ke dalam dua kategori berdasarkan fungsi. Kategori pertama terdiri atas peralatan yang melaksanakan operasi *input* dan *output*, kategori ini meliputi *keyboard*, *trackball*, *mouse*, *printer*, dan *display video*. Kategori kedua terdiri atas peralatan yang

diutamakan pada penyimpan data sekunder, yang mana penyimpan utamanya disediakan oleh memori utama komputer.Ada banyak sekali peralatan penyimpan, seperti *disk magnetic*, *optical disk*, *magnetic tapes*, yang mampu untuk menyimpan data yang besar. Peralatan yang umum digunakan pada Personal Computer (PC) yaitu *KeyBoard, Monitor, Mouse*, *Printer*, *Scanner*, *Disk Drive*, *Zip Drive*, dan *CD ROM Drive*.

2.10.4 CD-ROM

Mulai tahun 1983 sistem penyimpanan data di *optical disc* mulai diperkenalkan dengan diluncurkannya *Digital Audio Compact Disc*. Sejak saat itu mulai berkembanglah teknologi penyuimpanan pada *optical disc*. *CD-ROM* terbuat dari resin (*polycarbonate*) dan dilapisi permukaan yang sangat reflektif seperti alumunium. Informasi direkam secara digital sebagai lubang-lubang mikroskopis pada permukaan yang reflektif. Proses ini dilakukan degan menggunakan laser yang berintensitas tinggi. Permukaan yang berlubang ini kemudian dilapisi oleh lapisan bening. Informasi dibaca dengan menggunakan laser berintensitas rendah yang menyinari lapisan bening tersebut sementara motor memutar *disk*. Intensitas laser tersebut berubah setelah mengenai lubang-lubang tersebut kemudian terefleksikan dan dideteksi oleh fotosensor yang kemudian dikonversi menjadi data digital. Penulisan data pada *CD-ROM* hanya dapat dilakukan sekali saja. Walaupun demikian, optical disk ini memiliki keunggulan dari segi mobilitas. Bentuknya yang kecil dan tipis memudahkannya untuk dibawa-bawa. Kapasitas penyimpanannya pun cukup besar, yaitu 650

Mbytes. Sehingga media ini biasanya digunakan untuk menyimpan data-data sekali tulis saja, seperti *installer,* file lagu (mp3), ataupun data statik lainnya.

2.10.5 Mic (Microphone)

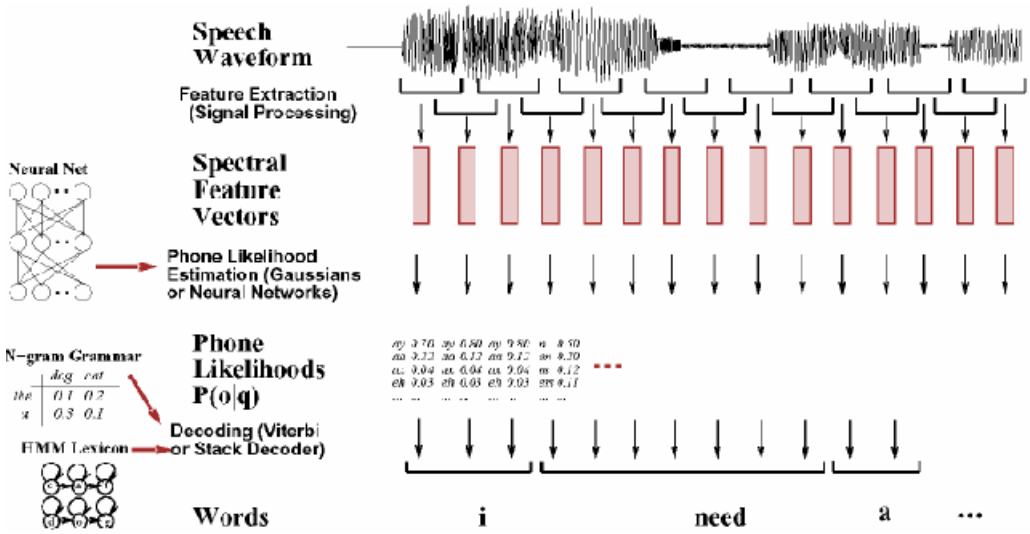
Kalau camera digunakan untuk memasukkan input berupa gambar (dan suara), maka mic digunakan hanya untuk memasukkan input berupa suara. Penggunaan mic tentu saja memerlukan perangkat keras tambahan untuk menerima input suara tersebut yaitu *sound card,* dan speaker untuk mendengarkan hasil rekaman.



Gambar 2.13 Microphone

2.11 Sistem Pengenalan Ucapan

Secara teknis, Sistem Pengenalan Ucapan dapat dijabarkan menggunakan contoh seperti pada Gambar berikut :



Gambar 2.14 Penjabaran Teknis Sistem Pengenalan Ucapan

Sumber sinyal yang berasal dari ucapan seseorang, dinyatakan sebagai *Speech Waveform* pada Gambar diatas. Sumber sinyal ini kemudian dicari ciri pentingnya yang dapat membedakan antara satu sinyal dengan sinyal lain. Pengungkapan ciri dilakukan dengan pengolahan atau ekstraksi sinyal.

Terdapat banyak cara dalam melakukan pengolahan sinyal untuk pengungkapan ciri, diantara yang terkenal adalah :

- *Mel Frequency Cepstral Coefficients* (MFCC)

  Analisis Transformasi Fourier menghasilkan frekuensi diskrit terhadap waktu dan dapat ditampilkan gambarnya. Biasanya nilai frekuensi dinyatakan menggunakan skala Mel, dimana skala skala ini linier untuk rentang dibagian yang rendah dan berbentuk logaritmik pada rentang bagian tinggi. Hal ini sesuai dengan karakteristik pendengaran manusia. Koefisien Sepstral frekuensi Mel inilah yang dijadikan ciri.

- *Linear Predictive Coding* (LPC)

  Cara ini menghasilkan koefisien dari persamaan linier yang memperkirakan nilai data ucapan beberapa saat terakhir. Vektor koefisien inilah yang dijadikan ciri.

Hasil dari pengolahan dan analisis sinyal adalah deretan vektor ciri ucapan, dinyatakan sebagai *spectral feature vector* pada Gambar diatas. Vektor-vektor ini dilatih menggunakan Jaringan Syaraf Tiruan. Pelatihan dimaksudkan agar pemrosesan vektor dapat dilakukan dengan cepat. Hasil dari pelatihan berupa deretan fonem-fonem yang membentuk data ucapan. Pemodelan fonem ini menggunakan model akustik *Hidden Markov Model* (HMM).

Proses pencarian kesamaan data fonem dilakukan dengan menggunakan Grammar N-Gram dan dikodekan kembali menggunakan algoritma analisis sinyal. Proses ini melihat nilai total penjajaran terbaik

2.12 Masalah-masalah yang Penting Pada Sistem Pengenalan Ucapan

Pada kenyataannya, Sistem Pengenalan Ucapan sampai saat ini masih perlu meningkatkan beberapa hal guna memperbaiki kinerjanya, terutama masalah akurasi. Dalam pengenalan ucapan disebutkan beberapa masalah mengapa sulit menigkatkan akurasi. Masalah tersebut dijelaskan sebagai berikut :

1. Semakin besar ukuran kosakata, semakin tinggi kesalahan dalam pengenalan ucapan. Sebagai contoh, 10 digit dari “nol” sampai “sembilan” dapat dengan mudah dikenali. Tetapi dengan meningkatnya ukuran *vocabulary* menjadi 200, 5000, atau 100000 maka bisa terjadi rata-rata

kesalahan 3%, 7%, atau 45% Sementara itu, setiap bahasa mempunyai huruf ketika diucapkan sulit untuk dibedakan. Sebagai contoh huruf C, D, J dalam kasus Bahasa Indonesia.

2. Tingkat ketergantungan kepada pembicara. Sistem pengenalan ucapan *Speaker Dependence* hanya ditujukan untuk pembicara tunggal. Sedangkan *Speaker Independence* ditujukan untuk siapapun. Jenis *Speaker Independence* sulit untuk dilakukan karena parameter suatu sistem pengenalan ucapan biasanya dilatih menggunakan *Speaker* tertentu, yang sangat spesifik ucapannya. Hasil penelitian menunjukkan bahwa *Speaker Independence* mempunyai kesalahan rata-rata 3 sampai 5 kali lebih besar dibanding *Speaker Dependence*.
3. Penggalan ucapan. *Isolated Speech* berarti ucapan kata tunggal. *Discontinuous Speech* berarti satu kalimat penuh tetapi kata dipilah berdasarkan keadaan diam (*silence*). *Continuous Speech* berarti ucapan kalimat secara alamiah. *Isolated* dan *Discontinuous Speech* relatif lebih mudah karena penggalan kata dapat dikenali dan kata per kata dapat diucapkan lebih “bersih”. Sedangkan *Continuous Speech* lebih sulit dilakukan karena tidak diketahui penggalan kata dan juga dipengaruhi oleh artikulasi gabungan kata.
4. Persyaratan bahasa. Berbagai pekerjaan sulit untuk dilakukan bilamana tidak memenuhi syarat-syarat penggunaan bahasa. Termasuk didalamnya adalah Sintaks, Simantik dan Grammarnya. Sebagai contoh, *query* terhadap jasa penerbangan dengan pernyataan ‘Pesawat itu berwarna

merah', diterima oleh Sintaks, Simantik maupun Grammar. Sedangkan pernyataan 'Pesawat itu sedang marah', diterima Sintaks dan Grammar tetapi ditolak oleh Simantik. Sehingga, semua struktur bahasa perlu ditinjau dalam membangun sistem pengenalan ucapan.

5. Ucapan Spontanitas. Sistem pengenalan ucapan dapat ditinjau dari pembicara yang sedang membaca skrip atau berbicara secara spontan. Pembicaraan spontan seperti emm, uh, e dan lain-lain, lebih sulit dikenali.
6. Kondisi lingkungan. Kinerja sistem pengenalan ucapan dapat juga diuji melalui lingkungan alamiah yang terdapat banyak interferensi, derau dan lain-lain. Sehingga usaha untuk dapat mengenali ucapan dari pembicara yang dimaksudkan dengan suara bersih telah banyak diteliti dengan berbagai metode. Diantaranya, penghilangan derau, penggandaan mikrofon, pembatasan lebar pita frekuensi, dan perubahan gaya ucapan.

Bahkan ada kelemahan Sistem Pengenalan Ucapan yang mendasar, yaitu bahasa pembicaraan akan lebih efektif bila digunakan antar manusia dan akan mempunyai banyak kekurangan bila diterapkan pada interaksi antara manusia dengan komputer. Hal ini diakibatkan oleh lambatnya komputer merepresentasikan informasi dan sulitnya melakukan *review* dan *edit*. Sehingga perancangan Sistem Pengenalan Ucapan harus mengerti tentang *acustic memory* dan *prosody*.