IMPLEMENTASI SISTEM PESAN VIA SUARA : KONVERSI SUARA KE TEKS PADA APLIKASI PENGIRIMAN PESAN BERBAHASA INDONESIA

IMPLEMENTATION OF VOICE ON MESSAGE : SPEECH TO TEXT ON INDONESIAN-LANGUAGE SENDED MESSAGE

Gede Kristian Wijaya Kusuma[1] **Astri Novianty, ST., MT[2]** **Andrew Brian Osmon, ST., MT[3]**

[1 2 3]Fakultas Teknik Elektro – Universitas Telkom
Jl. Telekomunikasi, Dayeuh Kolot Bandung 40257 Indonesia
[1]kristianwijayakusuma@gmail.com [2] astrinov@telkomuniversity.ac.id [3] abosmond@telkomuniversity.ac.id

ABSTRAK

Komunikasi adalah kebutuhan yang sangat penting dalam kehidupan manusia. Sebagai mahluk sosial, kebutuhan manusia akan komunikasi sangat tinggi. Dengan adanya teknologi telepon, pesan singkat, dan *e-mail* sangat membantu terwujudnya hubungan komunikasi, terutama komunikasi jarak jauh. Komunikasi jarak jauh sangat marak digunakan dalam berbagai urusan. Salah satunya dengan menggunakan teknologi pengiriman dan penerimaan pesan singkat atau *Short Message Service* (SMS). Penggunaan SMS dengan menulis dan membaca pesan banyak memiliki kekurangan khususnya disaat pengguna sedang tidak dapat mengoperasikan *smartphone*-nya. Contohnya saat berkendara, SMS yang diterima akan sulit untuk dibaca maupun dibalas. Untuk menanggulangi masalah tersebut dirancang suatu proyek besar yaitu Sistem Pesan Via Suara yang akan membantu menanggulangi masalah tersebut. Dalam Tugas Akhir ini, penulis akan membuat bentuk konversi suara ke teks pada proses pengiriman dengan menggunakan aplikasi pesan singkat yang merupakan salah satu bagian penting penyusun Sistem Pesan Via Suara tersebut. Pendeteksiannya menggunakan pengolahan sinyal suara dengan memanfaatkan ekstrasi ciri Mel-Scale Frequency Cepstral Coefficient dan klasifikasi K-Nearest Neighbor yang akan membantu mengkonversi sinyal suara ke teks dalam pengiriman pesan singkat dengan tepat.

Hasil yang didapatkan dari konversi suara ke teks pada pengiriman pesan adalah informasi dalam bentuk teks yang akan dikirim melalui pesan singkat. Keluaran sistem ini memiliki tingkat akurasi tertinggi secara realtime pada dependent speech dengan persentase 17,78% dengan jumlah tiga data latih. Sedangkan pada independent speech didapat akurasi tertinggi 11,11% dengan jumlah satu data latih.

Kata Kunci : SMS, *Mel-Scale Frequency Cepstral Coefficient, K-Nearest Neighbor, Speech to Text,* Android

ABSTRACT

Communication is a very important in human life's need. As social beings, humans need of communication are very high. With the phone technology, text message and e-mail helps the realization of the communication, especially long-distance communication. Long-distance communication is very widespread use in a variety of matters. One of that is by using the technology of sending and receiving short Message Service (SMS). The use of SMS by writing and reading messages has many disadvantaged, especially when the user is not able to operate his smartphone. For example, when user drive a vehicle, received SMS will be difficult to read and reply. To solve these problems, is designed a major project named Voice in Message System that will help resolve the problem. In this final project, the author built speech-to-text conversion on sended messages from the short message application. This system is one of the important constituent part of Voice in Message System. This conversion uses the Mel-Scale Frequency Cepstral Coefficient and K-Nearest Neighbour method that will help convert speech into text in the short message. The conversion of obtained result from speech to text is the information in the form of text that sended to the short message. The output of this system on realtime has the highest accuracy level dependent speech is 17.78 % with the number of three training data. While the independent speech obtained the highest accuracy of 11.11 % with the number of one training data.

Keywords: SMS, *Mel-Scale Frequency Cepstral Coefficient, K-Nearest Neighbor, Speech to Text,* Android

1. Pendahuluan

Sarana komunikasi merupakan salah satu penunjang utama dalam bekehidupan di dunia. Khususnya di Indonesia, pengguna sarana komunikasi pribadi sangatlah tinggi. Dimulai dari kalangan bawah yang berkehidupan cukup, hingga kalangan pejabat terbiasa menggunakan sarana komunikasi sebagai alat berkomunikasi. *Short Message Service* (SMS) adalah salah satu sarana komunikasi yang banyak digunakan karena tidak memakan banyak biaya, tidak menyita banyak waktu dan dapat berkomunikasi disela-sela kegiatan yang lainnya.

Akan tetapi, penggunaan SMS ini terbatas pada beberapa hal. Salah satunya pengguna dituntut untuk selalu dapat mengoperasikan perangkatnya saat menggunakan teknologi SMS ini. Pada saat-saat tertentu pengguna sulit mengoperasikan perangkatnya untuk mengakses informasi yang diterima melalui SMS ini. Contohnya saat berkendara, dimana pengguna tidak dapat menerima informasi yang diterima melalui SMS.

Kesulitan tersebut, terutama dalam pengiriman dan penerimaan SMS saat berkendara dapat diatasi dengan memberikan kemudahan dalam mengirim dan menerima informasi di SMS tersebut. Yaitu menggunakan teknologi *Speech*

To Text (STT) dan *Text To Speech (TTS)* dalam melakukan pengiriman dan penerimaan SMS. STT disistem ini digunakan untuk mengirim informasi pesan dalam bentuk teks dari hasil proses pengolahan sinyal suara yang direkam melalui microphone pada device.

2. Dasar Teori dan Perancangan

2.1 Sinyal Suara Manusia

Suara atau wicara, (*speech*) adalah bentuk komunikasi lisan manusia yang didasarkan pada kombinasi sintaksis leksikon dan nama yang diambil dari sejumlah besar kosakata. Setiap kata yang dilisankan tersusun atas kombinasi fonetis dari sejumlah kecil bunyi bahasa (vocal dan konsonan). Kosakata serta sintaksis dan bunyi bahasa penyusunnya ini membentuk ribuan bahasa yang pernah atau sedang digunakan manusia.[2]

Suara yang keluar dari mulut manusia akan memuat berbagai informasi seperti identitas pengucap, jenis gender, dialek, ekspresi, dan lain-lain. Proses produksi suara pada manusia dapat dibagi menjadi tiga buah proses fisiologis, yaitu: pembentukan aliran udara dari paru-paru, perubahan aliran udara dari paru-paru menjadi suara, baik voiced maupun unvoiced yang dikenal dengan istilah phonation, dan artikulasi yaitu proses modulasi/pengaturan suara menjadi bunyi yang spesifik. Faktor-faktor yang mempengaruhi produksi suara dapat dibagi menjadi dua bagian,yaitu laring dengan pita suaranya sebagai 'sumber suara' dimana suara dengan berbagai frekuensi yang kompleks dihasilkan dan 'filter' atau 'artikulator' dimana kita memodifikasi spektrum suara yang terbentuk tersebut dengan lidah, gigi, bibir, dan rongga mulut.[2]

Berdasarkan sinyal eksitasi yang dihasilkan pada proses produksi suara, sinyal suara ucapan dapat dibagi menjadi tiga bagian yaitu *silence, unvoiced,* dan *voiced* : [9]

1. Sinyal *silence* : sinyal pada saat tidak terjadi proses produksi suara ucapan, dan sinyal yang diterima oleh pendengar dianggap sebagai bising latar belakang.
2. Sinyal *unvoiced* : terjadi pada saat pita suara tidak bergetar, dimana sinyal eksitasi berupa sinyal random.
3. Sinyal *voiced* : terjadi jika pita suara bergetar, yaitu pada saat sinyal eksitasi berupa sinyal pulsa kuasi-periodik. Selama terjadinya sinyal voiced ini, pita suara bergetar pada frekuensi fundamental – inilah yang dikenal sebagai *pitch* dari suara tersebut.

2.2 Konsep Dasar Pengenalan Suara

Pengenalan Suara, (Speech Recognition) adalah proses identifikasi suara berdasarkan kata yang diucapkan dengan melakukan konversi sebuah sinyal akustik, yang ditangkap oleh perangkat input suara (*audio device*). Pengenalan suara juga merupakan sistem yang digunakan untuk mengenali perintah kata dari suara manusia dan kemudian diterjemahkan menjadi suatu data yang dimengerti oleh komputer. Pada saat ini, sistem ini digunakan untuk menggantikan peranan input dari keyboard dan mouse.[3]

Keuntungan dari sistem ini adalah pada kecepatan dan kemudahan dalam penggunaannya. Kata-kata yang ditangkap dan dikenali bisa jadi sebagai hasil akhir, untuk sebuah aplikasi seperti *command* dan *control*, penginputan data, dan persiapan dokumen. Parameter yang dibandingkan ialah tingkat penekanan suara yang kemudian akan dicocokkan dengan *template database* yang tersedia.[9]

2.3 Perbandingan Mode *Speaker-Dependent* dengan *Speaker-Independent* [9]

Mayoritas sistem pengenalan suara pada prinsipnya dapat digunakan pada mode *speaker-dependent* atau *speaker-independent* dan desain dari bagian-bagian sistem tersebut tergantung dari mode pelatihan (*training*) dari sistem. *Speaker-dependent* mengenali suara dengan data latih suara dari satu orang. Sistem ini mempelajari parameter dari karakteristik sinyal suara yang digunakan sebagai model dalam proses pengenalan suara. Sistem ini nantinya hanya digunakan untuk mengenali suara dari orang yang dijadikan sebagai data latih tersebut. Dengan demikian, pengenalan suara pada sistem ini akan menghasilkan hasil yang lebih akurat jika dibandingkan dengan sistem *independent-speaker*.

Sistem *independent-speaker* menggunakan suara data latih dari banyak orang dan digunakan untuk mengenali kata dari berbagai *speaker* yang memungkinkan sistem untuk dapat mengenali suara orang yang bukan sebagai suara data latih sistem. Meskipun sistem *dependent-speaker* memiliki tingkat akurasi yang lebih tinggi dibandingkan *independent-speaker*, terdapat suatu kelemahan pada sistem ini, yaitu apabila sistem ingin mengenali kata dari suara orang selain suara data latih, maka perlu dilakukan proses pelatihan (*training*) untuk mengenali suara orang baru tersebut.

2.4 Mel-Scale Frequency Cepstral Coefficient

Mel-Scale Frequency Cepstral Coefficient (MFCC) merupakan salah satu metode ekstraksi ciri sinyal suara yang berdasarkan prinsip karakteristik pendengaran telinga manusia. Kemampuan pendengaran pada telinga manusia tidak berskala linier namun dihitung dalam skala '*mel*' yang disebut sebagai '*Mel-Scale*'[4][5]. *Mel-Scale* merupakan skala yang

diambil berdasarkan pendekatan terhadap pendengaran manusia. Tahapan proses dalam Mel-Scale Frequency Cepstral Coefficient adalah : [4][5][9]

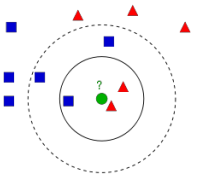
1. Frame Blocking
   Sinyal suara dibagi menjadi beberapa blok (atau *frame*, terdiri S sampel), yang digeser dari awal hingga akhir. Setiap *frame* dapat dianalisis dengan ukuran 0-20 ms sehingga tidak perlu menganalisis seluruh sinyal. Antara dua *frame* yang *adjacent* diterapkan proses *overlap* dengan nilai *overlap* yang sudah ditentukan. Tujuan dari *frame blocking* ini dilatarbelakangi bahwa kondisi sinyal audio yang berubah-ubah secara konstan, sehingga dilakukanlah proses *frame blocking* dengan panjang durasi yang lebih singkat dari sinyal aslinya dengan asumsi bahwa durasi yang singkat dari suatu sinyal suara kondisinya secara statistik tidak berubah-ubah.
2. Windowing
   Proses windowing dilakukan pada setiap *frame* dengan tujuan untuk meminimumkan diskontinuitas antar dua frame yang *adjacent*, khususnya pada bagian awal dan akhir. Adanya diskontinuitas menyebabkan terjadinya hilang informasi pada sinyal suara tersebut. Karena itu tahapan *windowing* perlu dilakukan untuk mendapatkan suatu sinyal secara keseluruhan tanpa kehilangan informasinya.
3. Fast Fourier Transform (FFT)
   Dasar melakukan Fast Fourier Transform ini untuk mentransformaikan sinyal dari domain waktu ke domain frekuensi (Hz). FFT adalah bentuk khusus dari persamaan integral *fourier*. Di dalam proses *Fast Fourier Transform* akan menghasilkan dua buah nilai yaitu nilai real dari respon frekuensi dan nilai imajiner dari respon *magnitude*. Output dari proses ini disebut dengan nama spektrum atau periodogram.
4. Mel-Frequency Wrapping
   Tahap ini merupakan proses pemfilteran dari spektrum setiap *frame* yang diperoleh dari tahapan sebelumnya, menggunakan sejumlah M filter segitiga dengan tinggi satu. Filter ini dibuat dengan mengikuti persepsi telinga manusia dalam menerima sinyal suara yang tidak linier dengan frekuensi suara. Oleh karena itu, untuk setiap nada dengan frekuensi f sebenarnya, diukur dalam Hz , *pitch* subjektif diukur pada skala yang disebut " *mel* ". Skala frekuensi *mel* adalah frekuensi linier yang berada dibawah 1000 Hz dan logaritmik untuk frekuensi diatas 1000Hz. Pada tahapan ini, dilakukan proses perkalian antara Mel-Spaced Filterbank dengan spektral daya dari periodogram yang telah didapatkan setelah proses FFT. Setelah dilakukan perkalian, hasil perkalian tersebut dijumlahkan.
5. Cepstrum
   Langkah terakhir adalah mengubah spektrum log-mel menjadi domain waktu. Hasil dari proses Cepstrum inilah yang disebut sebagai *Mel-Scale Frequency Cepstral Coeficient*. Representasi cepstral dari spektrum sinyal suara berasal dari sifat spektrum sinyal untuk analisa *frame* yang ada.

2.5 K-Nearest Neighbor

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan "jika-maka", berupa *decision tree*, *formula* matematis atau *neural network*.[6][7]

Algoritma K-NN adalah suatu metode yang menggunakan algoritma supervised. Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data. Tujuan dari algoritma K-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples*. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada nilai K. Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori.[6][7]

Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Jarak yang digunakan adalah jarak *Euclidean Distance*. Jarak Euclidean adalah jarak yang paling umum digunakan pada data numerik. Persamaan jarak Euclidean digunakan untuk mengukur kedekatan jarak (ciri) antara dua obyek, data latih dan data uji.
Data latih dengan jarak terdekat dikatakan sebagai tetangga (*Nearest Neighbor*) kemudian diurutkan dari jarak terdekat sampai terjauh. Tiap tetangga dapat berbeda satu sama lain ataupun sejenis. Tetangga sejenis dengan jumlah terbanyak di antara K tetangga terdekat adalah data latih yang sesuai dengan objek yang diklasifikasikan.



Gambar 2.1 Model metode KNN[8]

Pada gambar di atas dimisalkan K = 3, sehingga dapat dilihat dari 3 tetangga terdekat, data latih dengan ciri segi tiga merah memiliki jumlah paling banyak. Maka dari itu data uji (bulat hijau) dapat diklasifikasikan ke dalam data latih segi tiga merah. Nilai K tergantung uji coba pada data yang diperlukan. Secara umum, nilai K yang lebih besar memiliki akurasi yang lebih baik, namun membuat batas-batas antara setiap klasifikasi menjadi kurang jelas.[6][7][8]

2.6 Perancangan Sistem

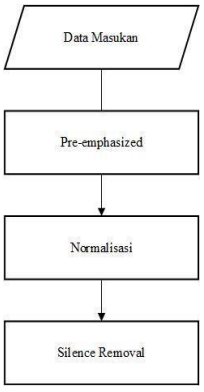
Dalam Tugas Akhir ini dirancang suatu sistem yang dapat mengkonversi sinyal suara menjadi informasi berbentuk teks pesan singkat. Sinyal suara yang direkam melalui microphone pada device akan dilakukan pemisahan kalimat menjadi kata. Setiap kata yang terekam akan dilakukan proses ekstraksi ciri kemudian mencocokkan dengan template database suara yang ada. Hasil dari proses matching akan didapatkan kata yang sesuai dengan perekam dan direpresentasikan ke dalam bentuk teks pada pesan singkat. Proses ekstraksi ciri menggunakan metode *Mel-Scale Frequency Cepstral Coefficient* (MFCC) dan proses klasifikasi menggunakan metode *K-Nearest Neighbor* (KNN). Untuk mendapatkan hasil kata yang sesuai rekaman dilakukan berbagai proses dari awal perekaman.

2.6.1 Tahap Akuisisi Data

Pada proses ini dilakukan proses perekaman suara berupa kata yang diucapkan dalam Bahasa Indonesia menggunakan perangkat lunak *Tape-a-Talk* serta perangkat keras berupa *device Smartphone* Lenovo A859. Suara yang direkam disimpan dalam format .wav dengan *sample rate* 8 KHz, *sample format* 16bit, dan tipe sinyal suara mono. Perbedaan tipe suara mono dan stereo ditunjukkan oleh jumlah kanalnya. Tipe suara mono terdiri dari satu kanal suara dan tipe suara stereo terdiri dari dua kanal suara.

2.6.2 Proses Pre-processing

Tahapan ini merupakan tahap awal pemrosesan sinyal suara yang terdiri dari tahapan *pre-emphasized*, kemudian proses *overlapping*, dilanjutkan dengan proses *windowing* serta terakhir proses *silent removal*. Adapun diagram blok untuk proses pre-processing ditunjukkan pada gambar dibawah:



Gambar 2.2 Activity diagram pre-processing

1. Pre-emphasized

Pre-emphasized adalah suatu teknik yang untuk penekanan kepada objek tertentu. Penekanan pada sinyal suara dilakukan pada frekuensinya, dimana sinyal suara yang direkam menggunakan perangkat lunak tidak terlepas dari pencampuran antara frekuensi tinggi dan frekuensi rendah. Teknik *pre-emphasized* akan memberikan penekanan pada frekuensi tinggi dan menghilangkan frekuensi rendah yang lemah. Hasil keluaran berupa suara yang terdengar lebih kecil seperti sengau tapi lebih baik dan jelas dibandingkan sinyal suara aslinya. Untuk teknik pre-emphasized digunakan filter FIR (*Finite Impluse Respons*).

2. Normalisasi

Pada proses normalisasi akan dilakukan perkalian amplitudo dari suatu sinyal suara dengan nilai amplitudo maksimum (yang diabsolut) dari sinyal suara tersebut. Proses ini dilakukan agar pada pemrosesan sinyal selanjutnya tidak dipengaruhi oleh amplitudo sinyal yang terlalu kecil atau amplitudo yang terlalu besar. Sehingga output dari proses normalisasi adalah sinyal yang memiliki rentang amplitudo antara -1 sampai +1.

3. Silence Removal

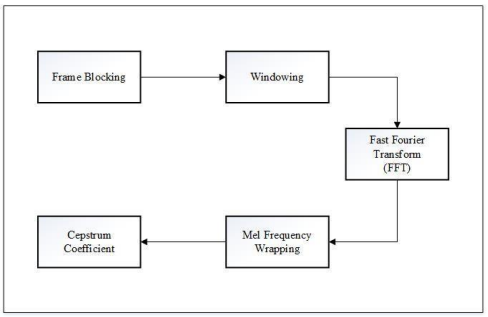
Proses *silence removal* dilakukan untuk menghilangkan daerah *silence* dari sinyal suara untuk meningkatkan akurasi sistem. Proses ini dilakukan dengan mencari nilai standar deviasi dari sinyal suara. Nilai standar deviasi tersebut dihitung berdasarkan rumus berikut : [9]

$$std = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{1/2}$$

Nilai sinyal yang lebih kecil dari nilai standar deviasi sinyal suara akan dianggap sebagai daerah *silent*. Setelah didapatkan daerah *silent*, dicari indeks maksimum dan minimum dari daerah *silence* tersebut dengan tujuan memotong daerah *silent* dari indeks minimum sampai daerah maksimum tersebut, dengan demikian, daerah *silent* yang dipotong adalah daerah awal dan akhir dari sinyal suara. Daerah *silent* yang berada ditengah-tengah sinyal suara tidak akan dianggap sebagai daerah *silent*

2.6.3 Proses Ekstraksi Ciri

Pada penelitian kali ini, akan dilakukan ekstraksi ciri pada sinyal bicara menggunakan metode Mel-Scale Frequency Cepstral Coefficient (MFCC). Output dari proses ini adalah nilai MFCC sebanyak jumlah koefisien yang kita inginkan. Apabila kita menginginkan nilai MFCC sebanyak 12 koefisien, maka masing-masing frame dari sinyal suara akan direperesentasikan ke dalam 12 nilai MFCC. Maka ketika ukuran matriks sinyal yang telah di framing adalah 100 x 2000, maka hasil keluaran proses ekstraksi ciri adalah matriks berukuran 12 x 2000. Pada proses ekstraksi ciri terdapat beberapa tahapan, yaitu:



Gambar 2.3 Proses Ekstraksi Ciri (MFCC).

2.6.4 Proses Deteksi

Pada proses deteksi dilakukan dengan membandingkan jumlah koefisien MFCC yang telah didapat dari masing-masing sampel suara. Setiap sampel tidak memiliki jumlah nilai MFCC yang sama karena setiap sinyal suara memiliki jumlah *frame* yang berbeda. Untuk mendapatkan objek uji yang sesuai dengan objek *training* maka setiap nilai MFFC tersebut dibandingkan ke semua sampel data training. Hasil dari proses ini akan didapat nilai jarak dari objek uji terhadap semua objek *training* dan objek *training* yang memiliki nilai terdekat dengan nilai objek uji akan menjadi objek hasil uji dengan jumlah tetangga yang sudah ditentukan pada proses pengujian, yaitu menggunakan nilai k=1 dan k=3.

2.7 Perancangan Aplikasi

Perancangan aplikasi "VoiceOnMessage" menggunakan pemodelan *Unified Modelling Language* (UML) yang terdiri dari *use case diagram, class diagram* dan *activity diagram.*

2.8 Parameter Performansi Sistem

Tujuan dari pembuatan analisis sistem adalah untuk mengetahui performansi sistem, terutama dari sisi akurasi sitem. Performansi sistem dapat diukur salah satunya berdasarkan parameter berikut ini:

Performansi sistem dinilai berdasarkan akurasi sistem dengan menghitung persentase jumlah kata yang dideteksi dengan benar,persamaan untuk perhitungan akurasi yaitu:

$$\text{Akurasi} = \frac{\text{Jumlah Benar}}{\text{Jumlah Total}} \times 100\%$$

Untuk perhitungan persamaan kesalahan pembelahan kalimat, yaitu :

$$\text{Akurasi Kesalahan} = \frac{\text{Jumlah Salah}}{\text{Jumlah Total}} \times 100\%$$

3. ANALISIS DAN KELUARAN SISTEM

Untuk mengetahui performansi aplikasi yang telah dirancang, maka perlu dilakukan pengujian terhadap sistem yang telah dikembangkan. Pada tahap pengujian kehandalan program yang telah dibuat untuk mengkonversi sinyal suara menjadi informasi teks. Tingkat keberhasilan sistem ditunjukkan dengan tingkat akurasi sistem yang dihasilkan dari kombinasi berbagai parameter pengujian.

3.1 Pengujian Sistem

Untuk mengetahui performansi sistem dilakukan beberapa skenario pengujian pada aplikasi yang telah dirancang, scenario yang dilakukan meliputi :

3.1.1 Skenario 1 – Pengujian Dependent Speech

Tujuan dari pengujian ini adalah mengetahui apakah sistem dapat mendeteksi kata dengan benar dari hasil data masukan berupa sinyal suara yang diambil dari satu orang. Pengujian dilakukan dengan menggunakan data latih yang sama dengan data uji dengan jumlah data latih sebanyak satu data dan tiga data dan dengan batasan 90 kata.

Dari seluruh hasil pengujian scenario pertama, didapatkan akurasi untuk sistem dengan jumlah data latih satu yaitu sebesar 14,44% sedangkan untuk jumlah 3 data latih didapat akurasi sebesar 17,78%**.**

3.2.2 Skenario 2 – Pengujian Independent Speech

Tujuan dari pengujian ini adalah mengetahui apakah sistem dapat mendeteksi kata dengan benar dari hasil data masukan berupa sinyal suara yang diambil dari gabungan suara laki-laki dan perempuan.. Pengujian dilakukan dengan menggunakan enam orang pembica yang terdiri dari tiga pembicara laki-laki dan tiga orang pembicara permpuan. Sistem diuji dengan batasan 90 kata.

Dari seluruh hasil pengujian scenario kedua, didapatkan akurasi untuk sistem deteksi kata adalah paling tertinggi 11,11% dengan jumlah satu data latih sedangkan bertambahnya jumlah data latih semakin menurunkan performansi sistem dimana akurasi 8,89% didapat dengan jumlah sepuluh data latih, akurasi 8,89% didapat dari jumlah limabelas data latih dan akurasi terkecil 4,44% dari tiga puluh data latih.

3.2.3 Skenario 3 – Pengujian Pembelahan Kalimat

Pada pengujian skenario ketiga ini akan dianalisi performansi sistem dalam membelah kalimat menjadi kata-kata. Pengujian dilakukan dengan sampel kalimat dari tiga kata, lima kata dan tujuh kata yang masing-masing kata dipilih secara acak dari batasan kata yang ada. Pemilihan kata juga didasarkan pada ada tidaknya daerah silence pada pertengahan kata karena apabila terdapat daerah silence dipertengahan kata tentu suatu kata akan dianggap dapat terbelah menjadi dua buah kata atau lebih. Hasil yang didapat dari pembelahan kalimat adalah akurasi kesalahan tertinggi yaitu 66,67% pada kalimat dengan tiga kata, sedangkan akurasi kesalahan terkecil didapat 28,57% pada kalimat dengan tujuh kata.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil rancangan dan pengujian pada bab-bab sebelumnya dapat ditarik beberapa kesimpulan sebagai berikut.

a. Sinyal suara yang masuk berhasil diekstak ciri melalu metode MFCC dan mendapatkan nilai ciri. Hasil ekstraksi kemudian diklasifikasikan dengan menggunakan metode KNN dengan cara menentukan hasil terbaik berdasarkan jarak terdekat (tetangga) objek uji terhadapa data latih.

b. Hasil rancangan metode ekstraksi ciri dari MFCC dan klasifikasi KNN berhasil diimplementasikan kedalam bentuk aplikasi berbasis Android di mobile smartphone. Hasil data masukan berupa suara dari microphone dapat langsung di proses untuk menghasilan informasi teks yang sesuai dengan data masukan suara dari microphone.
c. Hasil pengujian kinerja menunjukkan akurasi tertinggi sistem untuk dependent speech adalah 17,78% dengan tiga data latih dan 14,44% untuk jumlah data latih satu data. Untuk pengujian independent speech dengan data latih gabungan didapat akurasi tertinggi dengan jumlah data latih satu yaitu sebesar 11,11%, akurasi 8,89% untuk jumlah data latih sepuluh data, akurasi 8,89% untuk jumlah data latih sebanyak 15 data dan yang terakhir dengan data latih sebanyak 30 data latih didapat akurasi sebesar 4,44%. Untuk pengujian pembelahan kalimat didapat akurasi kesalahan pembelahan terbesar adalah 66,67% sedangkan akurasi kesalahan pembelahan terkecil yaitu 28,57% dari kalimat dengan tujuh kata.

4.2 Saran

Adapun saran untuk pengembangan penelitian ini sebagai berikut.

a. Mencoba menggunakan metode ekstraksi ciri lain, seperti LPC (Linier Predictive Coding) untuk membantu perbaikan akurasi klasifikasi KNN dan penambahan metode pre-processing lain untuk membedakan nada, intonasi suara. Untuk perbaikan kualitas suara digunakan algoritma decision directed, algoritma two step noise reduction dan algoritma harmonic regeneration noise reduction.
b. Untuk proses deteksi sebaiknya digunakan metode yang lebih baik dalam pengambilan keputusan. Algortima Genetika dan Hidden Markov Model adalah beberapa algoritma yang dapat memberikan keputusan yang lebih baik sehingga didapat hasil deteksi yang lebih baik. Untuk hasil akurasi yang lebih baik juga bisa digunakan beberapa algoritma clustering untuk pengelompokkan nilai ciri sehingga tidak ada nilai ciri data latih yang menganggu nilai ciri data latih lainnya karena kemiripan nilai.
c. Menggunakan bahasa pemrograman lain, seperti C/C++, untuk pengembangan aplikasi pengolahan suara secara real-time.

DAFTAR PUSTAKA

[1] Safaat. H, Nazruddin (2012). Pemrograman Aplikasi Mobile Smartphone dan Tablet PC Berbasis Android. Informatika.
[2] Fadlisyah, Bustami, & M.Ikhwanus (2013). Pengolahan Suara. GRAHA ILMU.
[3] Mutohar, Amin. (2007). Voice Recognition Diakses 17 November 2014, dari http://mutohar.files.wordpress.com/2007/11/voice-recognition.pdf
[4] Ittichaichareon, Chadawan, Siwat Suksri, dan T.Yingthawornsuk. (28 Juli 2012). Speech Recognition Using MFCC. Diakses 9 Desember 2014, dari http://psrcentre.org/images/extraimages/712576.pdf
[5] Dhingra, Shivanker.D, Geeta Nijhawan, dan Poonam Pandit. (8 Agustus 2013). Isolated Speech Recognition Using MFCC And DTW. Diakses 9 Desember 2014, dari http://www.ijareeie.com/upload/2013/august/20P_ISOLATED.pdf
[6] Nugraheni, Y. (6 April 2013). Algoritma KNN. Diakses 30 Oktober 2013, dari http://yohananugraheni.files.wordpress.com/2013/04/4_knn.pptx
[7] Riska.Y, Marji, dan Dian Eka.R. (2014). Klasifikasi Suara Berdasarkan Gender (Jenis Kelamin) Dengan Metode K-Nearest Neighbor (KNN). Diakses 20 Agustus 2014, dari http://http://www.academia.edu/7699117/KLASIFIKASI_SUARA_BERDASARKAN_GENDER_JENIS_KELAMIN_DENGAN_METODE_K-NEAREST_NEIGHBOR_KNN
[8] Mulyana, Baskara. (2014).Identifikasi Jenis Bunga Anggrek Menggunakan Pengolahan Citra Digital Dengan Metode KNN. Bandung : Universitas Telkom.
[9] Rahayu, Intan. (2014). Analisis dan Simulasi Sistem Penerjemah Kata Berbahasa Bali ke Bahasa Inggris Berbasis Speech To Text Secara Real Time Menggunakan Metode Klasifikasi Hidden Markov Model. Bandung : Universitas Telkom.