IMPLEMENTASI TEKNIK *DYNAMIC TIME WARPING* (DTW) PADA APLIKASI *SPEECH TO TEXT*

Candra Dinata[1], Diyah Puspitaningrum[2], Ernawati[3]

[1,2,3]Program Studi Teknik Infomatika, Fakultas Teknik, Universitas Bengkulu.
Jl. WR. Supratman Kandang Limun Bengkulu 38371A INDONESIA
(telp: 0736-341022; fax: 0736-341022)
[1]candra07dinata06@gmail.com, [2]diyah.puspitaningrum@unib.ac.id, [3]ernawati@unib.ac.id

ABSTRAK

Suara/ucapan adalah salah satu cara kita sebagai manusia untuk berkomunikasi dan mengekspresikan diri. *Speech to text* (ucapan ke text), merupakan salah satu bidang sains computer yaitu bidang pengolahan suara. *Speech to text* (STT) adalah penerjemahan kalimat (kata yang diucapkan) ke dalam *text*. STT merupakan proses pengolahan suatu sinyal suara, mengekstrak fitur dari sinyal suara tersebut yang selanjutkan dibandingkan dengan hasil ekstraksi dari sinyal suara yang lain untuk dapat dikenali persamaannya. Penelitian ini merancang dan membangun suatu program aplikasi *Speech to Text* yang mampu identifikasi suatu sinyal suara menggunakan perangkat lunak simulasi MATLAB R2016a. Terdapat dua proses umum pada bidang pengolahan suara, yaitu ekstraksi fitur dan pencocokan fitur. Pada sistem ini metode *mel-frequency cepstral coefficients* digunakan untuk mengekstraksi fitur dan metode *dynamic time warping* digunakan untuk pencocokan fitur. Metode DTW yang digunakan dapat menghitung jarak atau selisih antara dua data yang dibandingkan. Rata-rata akurasi yang didapat setelah dilakukan percobaan pada pengujian kata adalah 95.85% dan pada pengujian kalimat adalah 94%.

Kata Kunci: *Pengolahan Suara, Speech to Text, MFCC, DTW*

ABSTRACT

Voice / speech is one of the ways we as human beings to communicate and express themselves. Speech to text (STT), is one of computer science is the field of sound processing. Speech to text (STT) is the translation of the sentence (the spoken word) in the text. STT is a voice signal processing, extracting features from the speech signal and then compared it with the extraction of the other sound signal to recognize the signal similarities. This research design and build an application program Speech to Text that is capable of identifying a sound signal using simulation software MATLAB R2016a. There are two common processes in the field of sound processing, feature extraction and matching features. In this system, the method mel-frequency cepstral coefficients are used to extract features and dynamic time warping method used for matching features. DTW method used can calculate the distance or the difference between the two data being compared. The average accuracy is obtained after experiments on the test word was 95.85% and the testing of the sentence is 94%.

Keywords: *Voice Processing, Speech to Text, MFCC, DTW*

I. PENDAHULUAN

Suara/ucapan adalah cara kita sebagai manusia untuk berkomunikasi dan mengekspresikan diri. Secara perlahan dan dengan didukung oleh perkembangan teknologi ,kebutuhan akan adanya sistem dan aplikasi yang mampu menganalisis dan mengidentifikasi suatu sinyal suara pun semakin tinggi. Pemanfaatan aplikasi ini juga semakin berkembang, mulai dari sarana pembelajaran hingga bidang keamanan. Dalam bidang pembelajaran misalnya, kemampuan menguasai bahasa asing adalah salah satu syarat untuk dapat bergabung dalam komunitas masyarakat yang lebih luas. Sarana untuk mempelajari hal tersebut juga semakin banyak, salah satunya adalah menggunakan aplikasi *Speech To Text.*

Speech to text (ucapan ke *text*), biasanya merujuk ke pengolahan sinyal suara *(Speech Recognition)*. Menurut [1] *Speech Recognition* (SR) dalam bidang sains computer, adalah penerjemahan kalimat (kata yang diucapkan) ke dalam text. Juga dikenal sebagai "*automatic speech recognition*", "ASR", "STT", "*speech to text*", atau hanya "*computer speech recognition*".

Ide mengenai interaksi manusia dan mesin mendorong penelitian dalam pengenalan suara. ASR menggunakan proses dan tekologi terkait untuk mengubah sinyal ucapan menjadi rangkaian kata atau unit linguistik lain dengan cara algoritma diimplementasikan sebagai program komputer. Sistem pemahaman ucapan saat ini mampu memahami masukan suara untuk ribuan kosakata di lingkungan operasional [2].

Sistem pengenalan suara dapat diklasifikasikan menjadi tiga tipe: (1) Suara terisolir; (2) Suara diskontinu; (3) Suara kontinu [3]. Dalam penelitian ini, sistem pengenalan suara yang digunakan yaitu suara terisolir. Menurut [3] pengenalan suara terisolir mengenali satu kata dari pengguna, setiap kata yang terucap memiliki informasi dalam bentuk sinyal suara. Adapun algoritma yang akan digunakan adalah *Dynamic Time Warping* (DTW). DTW adalah algoritma yang digunakan untuk mengukur kesamaan antara dua sekuens yang mungkin berbeda dalam waktu atau kecepatan. Dua ucapan dari kata yang sama oleh pengguna yang sama dapat memiliki waktu yang berbeda. Sebagai contoh, *two* dapat dilafalkan dengan *to* atau *too* [4].

Waktu keselarasan ucapan yang berbeda adalah masalah inti untuk pengukuran jarak dalan pengenalan ucapan. Pergeseran kecil mengakibatkan identifikasi yang salah. *Dynamic Time Warping* adalah metode yang efisien untuk memecahkan masalah keselarasan waktu. Algoritma DTW ditujukan untuk menyelaraskan dua sekuen vector dengan membelokkan sumbu waktu berulang-ulang sampai kecocokan optimal antara dua sekuen ditemukan. Algoritma ini melakukan sebagian pemetaan linear dari sumbu axis untuk menyelaraskan kedua sinyal [4].

Berdasarkan permasalahan yang ada dan analisis metode yang telah dipaparkan, maka dari itu akan dilakukan penelitian dengan judul "Implementasi/penerapan teknik *(Dynamic Time Warping)* DTW ke aplikasi *Speech To Text".*

II. TINJAUAN PUSTAKA

2.1 Pengenalan *Digital Signal Processing* (DSP)

Dunia ilmu pengetahuan dan teknik diisi dengan sinyal: citra dari pesawat antariksa jarak jauh, tegangan yang dihasilkan oleh jantung dan otak, radar, dan sonar gema, getaran seismic, dan aplikasi lainnya yang tak terhitung jumlahnya. *Digital Signal Processing* adalah ilmu yang menggunakan komputer untuk memahami jenis data ini. Hal ini mencakup berbagai macam tujuan : penyaringan, pengenalan suara, peningkatan citra, kompresi data, jaringan syaraf, dan banyak lagi. DSP adalah salah satu teknologi yang paling kuat yang membentuk ilmu pengetahuan dan teknik dalam abad kedua puluh satu [5]. Tema yang diangkat dalam penelitian ini adalah pengenalan suara, sehingga keseluruhan materi dari DSP tidak diuraikan.

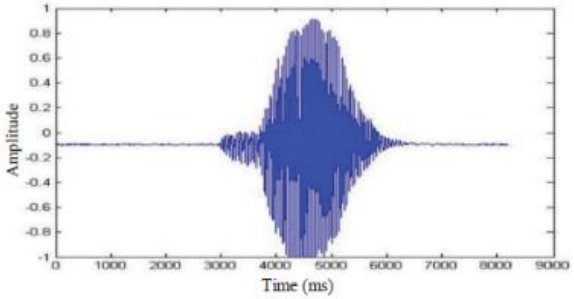
2.2 Sistem Pengenalan Suara

1. Pengenalan Suara Terisolir

Pengenalan suara terisolir adalah untuk mengenali satu kata dari pengguna dan merupakan pendekatan relatif yang mudah. Setiap kata yang terucap memiliki informasi dalam bentuk sinyal suara. Gambar 2.1 merupakan contoh sinyal suara di MatLab. Itu adalah kata *'No'* disajikan dalam domain waktu [3].

Sinyal suara terdiri dari : suara yang berguna dan "kebisingan". Untuk memisahkan dua bagian dari ucapan manusia terserbut,

digunakan metode pemisahan sinyal *cepstrum* yang efisien. Dalam suatu penelitian, [4] menggunakan MFCC (Mel Frecuency Cepstrum Coefficients) dan DTW (Dynamic Time Warping) untuk mengenali suara terisolir. Algoritma DTW adalah pendekatan untuk menghitung kesamaan antara dua deret waktu yang mungkin berbeda dalam waktu atau kecepatan [3].



Gambar 2.1 contoh kata 'No' di Matlab [3]

2. Suara Diskontinu

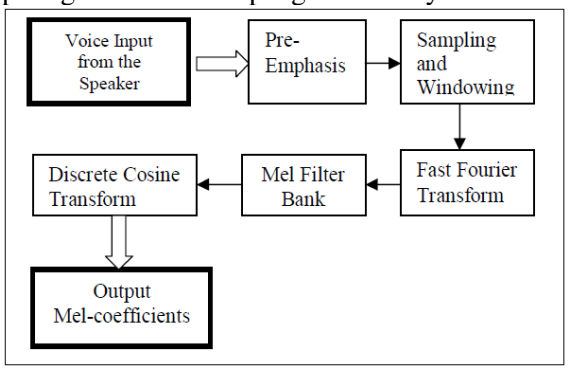
Suara diskontinu memiliki kemiripan dengan suara terisolir. Dimana terdapat jeda yang disengaja antara kalimat dan mungkin lebih dari satu orang dalam percakapan. Dhingra [6] mengajukan metodologi baru untuk menganalisa suara diskontinu dalam percakapan Uni Eropa. Metodologi ini mengidentifikasi perubahan topik sesuai dengan intervensi antar kalimat dan juga ketika pembicara berganti. Dan berdasarkan perubahan topik, fluktuasi diskursif yang tanpa henti dapat dibedakan. Ini adalah perbedaan utama dibandingkan dengan metode lain yang semua elemen diterjemahkan [3].

3. Suara Kontinu

Suara kontinu atau berkesinambungan mewakili performa alami dari orang bicara. Dimana tidak ada jeda yang disengaja. Itu adalah situasi tersulit untuk mengenali pembicaraan. Teknologi di bidang ini belum dikembangkan dengan baik dan metode yang ada jelas kurang dalam keakuratan [3].

2.3 Metode *Mel Frequency Cepstral Coefficients* (MFCC)

Dalam proses pengenalan sinyal suara dibutuhkan suatu metode yang digunakan untuk mengekstrak sinyal suara tersebut terlebih dahulu. Hasil dari ekstraksi sinyal suara tersebut selanjutnya akan digunakan untuk tahap pencocokan sinyal yang menggunakan metode *dynamic time warping*. Pada penelitian ini proses ekstraksi sinyal suara menggunakan metode mel frekuensi cepstral koefisien. Algoritma ini dapat menghitung koefisien unik untuk sampel tertentu. Kesederhanaan prosedur untuk pelaksanaan MFCC membuat teknik ini paling disukai untuk pengolahan sinyal suara.



Gambar 2.2 Blok diagram untuk memperoleh MFC koefisien [7]

2.4 Metode *Dynamic Time Warping*

Dynamic Time Warping (DTW) adalah sebuah algoritma yang menghitung jalur pembengkokan optimal antara dua deretan waktu. Algoritma menghitung kedua nilai jalur pembengkokan antara dua deretan dan jarak di antara mereka. Algoritma ini dimulai dengan perhitungan jarak lokal antara elemen dari dua urutan menggunakan berbagai jenis jarak. Metode yang paling sering digunakan untuk perhitungan jarak adalah jarak absolut antara nilai-nilai dari dua elemen (jarak euclidean) [8].

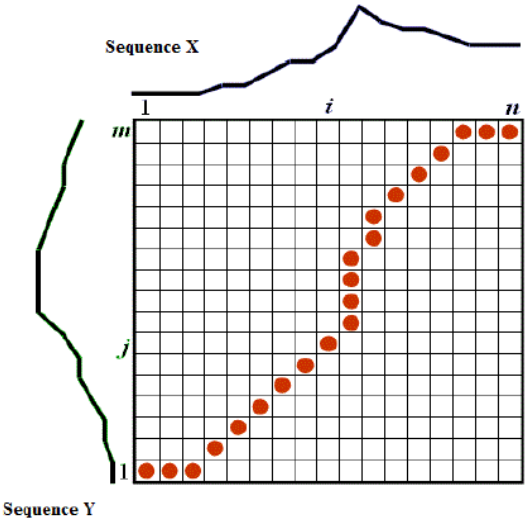
Dua ucapan dari kata yang sama oleh pengguna yang sama dapat memiliki waktu yang berbeda. Sebagai contoh, *two* dapat dilafalkan dengan *to* atau *too.* DTW menyelesaikan masalah ini dengan menyelaraskan kata-kata dengan benar dan menghitung jarak minimum antara dua kata. Sebuah matriks jarak lokal dibentuk untuk semua segmen dalam kata sampel dan *template* kata [4].

Waktu keselarasan ucapan yang berbeda adalah masalah inti untuk pengukuran jarak dalan pengenalan ucapan. Pergeseran kecil mengakibatkan identifikasi yang salah. *Dynamic Time Warping* adalah metode yang efisien untuk memecahkan masalah keselarasan waktu. Algoritma DTW ditujukan untuk menyelaraskan dua sekuen vector dengan membelokkan sumbu waktu berulang-ulang sampai kecocokan optimal antara dua sekuen ditemukan. algoritma ini melakukan sebagian pemetaan linear dari sumbu axis untuk menyelaraskan kedua sinyal.

Anggap saja terdapat dua urutan *vector* di ruang *n*-dimensi [4].:

$x = [x_1, x_2, ..., x_n]$ dan $y = [y_1, y_2, ..., y_n]$

Dua urutan tersebut selaras di sisi kotak, dengan satu di atas dan lainnya di sisi kiri. Kedua urutan mulai di bagian bawah kiri grid.



Gambar 2.3 Jarak Grid Global [4]

Dalam setiap sel, ukuran jarak ditempatkan, membandingkan unsur-unsur yang sesuai dari dua sekuens. Jarak antara dua titik dihitung melalui jarak Euclidean [4].

Dist (x, y) = |x - y| = [$(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2$]$^{1/2}$

Pencocokan terbaik atau keselarasan antara dua sekuens ini adalah jalan melalui grid, yang meminimalkan total jarak antara mereka, yang disebut sebagai jarak global. Keseluruhan jarak (jarak Global) dihitung dengan menemukan dan pergi melalui semua rute yang mungkin melalui grid, masing-masing menghitung jarak keseluruhan.

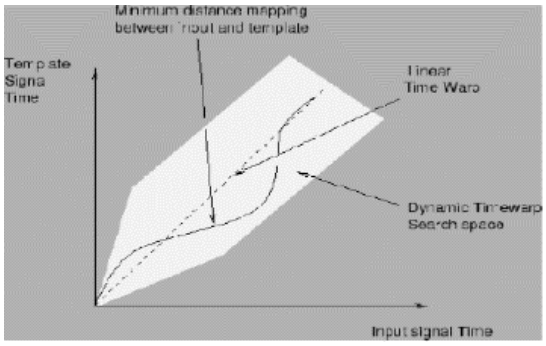
Jarak global minimum dari jumlah jarak (jarak Euclidean) antara unsur-unsur individual di jalan dibagi dengan jumlah dari fungsi pembobotan. Untuk setiap urutan cukup panjang jumlah kemungkinan jalan melalui grid akan sangat besar. mengukur jarak global diperoleh dengan menggunakan rumus rekursif [4].

$GD_{xy} = LD_{xy} + \min (GD_{x-1\ y-1}, GD_{x-1\ y}, GD_{x\ y-1})$

Dimana,

GD = Global Distance (*overall distance*)

LD = Local Distance (*Euclidean distance*)



Gambar 2.4 Dynamic Time Warping [4]

III. METODE PENELITIAN

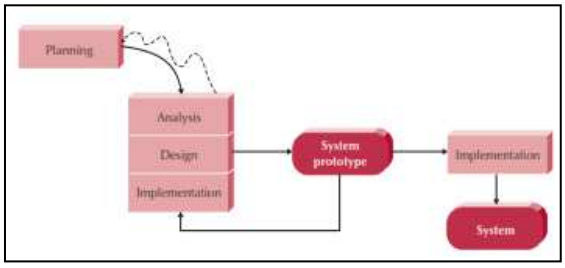
3.1 Dataset Penelitian

Data set penelitian dengan sampel 250 data kalimat suara dan 250 data kata suara yang merupakan suara dari 5 orang dewasa, 3 suara laki-laki dan 2 suara perempuan. Jumlah sampel yang diambil dari masing-masing orang yaitu 50 data kalimat dan 50 data kata data audio, dimana 40 data kalimat digunakan sebagai data latih,10 data kalimat untuk data uji, dan 50 data suara sebagai data uji.

Satu data kalimat suara terdiri dari 5 kosakata. Setiap orang melakukan perekaman 10 kalimat yang diulang sebanyak 5 kali dan semua kata yang terkandung di kalimat tersebut sebanyak 1 kali. *Training* data akan memisahkan kata-kata yang ada di dalam suatu kalimat, sehingga menghasilkan 5 data latih. Penyimpanan dan pengujian data dilakukan dalam kondisi yang sama yaitu, ruangan kecil dan tertutup dengan sedikit gangguan suara.

3.2 Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan adalah dengan pendekatan *prototyping*. *Prototyping* mengerjakan fase analisis, perancangan, dan implementasi secara bersamaan pengembangan yang cepat dan pengujian terhadap model kerja (*prototype)* dari aplikasi baru melalui proses interaksi dan berulang-ulang yang biasa digunakan ahli sistem informasi.



Gambar 2.3 Diagram Alir Metodologi *Prototyping* [9]

3.3 Metode Pengujian Sistem

1. *White Box Testing*

Pada *white box testing* akan digunakan metode pengujian *basis path*, yaitu salah satu teknik pengujian *white box* yang diusulkan pertama kali oleh Tom McCabe. Metode basis ini memungkinkan desainer *test case* mengukur kompleksitas logis dari desain prosedural dan menggunakannya sebagai pedoman untuk menetapkan basis set dari jalur eksekusi. *Test case* yang dilakukan untuk menggunakan *basis set* tersebut dijamin menggunakan setiap *statement* di dalam program paling tidak sekali selama pengujian.

2. *Black Box Testing*

Teknik pengujian *black box* yang dilakukan pada penelitian ini adalah teknik *equivalence partitioning*, yaitu teknik pengujian yang membagi domain *input* dari suatu program ke dalam kelas data, menentukan kasus pengujian dengan mengungkapkan kelas-kelas kesalahan

3.4 Uji Kelayakan Sistem

Terdapat dua jenis rasio kesalahan pencocokkan, yaitu rasio kesalahan kecocokan (*false match rate)* dan rasio kesalahan ketidak cocokkan (*false non match rate*).

a. Rasio Kesalahan Kecocokan

False Match Rate (FMR) menyatakan probababilitas sampel dari pengguna cocok dengan acuan yang diambil secara acak milik pengguna yang berbeda. FMR disebut juga *false posititive.* Rasio kesalahan kecocokan dihitung dengan rumus [10] berikut.

Persentase Kesalahan Kecocokan = $\frac{\sum Data\ yang\ cocok}{\sum jumlah\ data\ input} \times 100\%$.....…(1)

b. Rasio Kesalahan Ketidakcocokan

False Non Match Rate (FMR) menyatakan probabilitas sampel dari pengguna tidak cocok dengan acuan lain yang diberikan pengguna yang sama. FNMR disebut juga *false Negative*. Rasio kesalahan ketidak cocokan dihitung dengan rumus [10] berikut.

Persentase Kesalahan Ketidakcocokan = $\frac{\sum Data\ yang\ cocok}{\sum jumlah\ data\ input} \times 100\%$....(2)

IV. ANALISIS DAN PERANCANGAN SISTEM

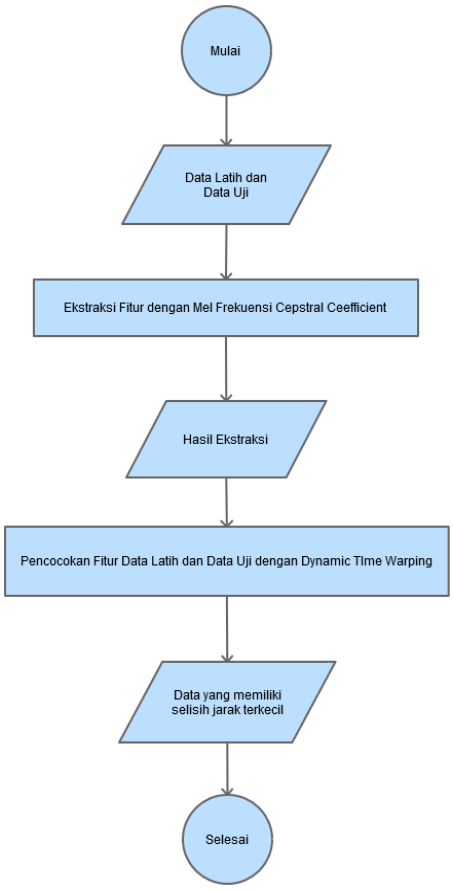
4.1 Jenis Penelitian

Penelitian murni adalah penelitian yang diperuntukan bagi pengembangan suatu ilmu pengetahuan serta diarahkan pada pengembangan teori-teori yang ada atau menemukan teori baru. Peneliti yang melakukan penelitian dasar memiliki tujuan mengembangkan ilmu pengetahuan tanpa memikirkan pemanfaatan secara langsung dari hasil penelitian tersebut [11]. Sedangkan penelitian terapan dilakukan berkenaan dengan kenyataan-kenyataan praktis, penerapan, dan pengembangan ilmu pengetahuan yang dihasilkan oleh penelitian dasar dalam kehidupan nyata. Penelitian terapan berfungsi untuk mencari solusi tentang masalah-masalah tertentu [12].

Berdasarkan definisi kedua jenis penelitian diatas yang dilihat dari tujuannya, maka penelitian yang diajukan oleh penulis dalam judul implementasi teknik *Dynamic Time Warping (*DTW) pada aplikasi *Speech To Text* ini termasuk dalam penelitian terapan (*applied research).* Dalam penelitian ini, penulis berusaha untuk menerapkan algoritma DTW pada suatu sistem aplikasi sehingga dapat mengidentifikasi suatu sinyal suara dan menampilkan keterangan mengenai sinyal suara tersebut.

4.2 Analisis Alur Kerja Sistem

Alir sistem merupakan hasil analisis perancangan tahapan kerja sistem yang akan dibangun. Alur ini dimulai dari user memasukkan *(input)* data sampai menghasilkan keluaran *(output).* Dalam sistem ini, *input*-nya berupa suara orang dewasa dan informasi data suara tersebut. Sedangkan untuk *output*-nya adalah hasil identifikasi suara dalam bentuk teks. Gambar 4.1 merupakan aliran data yang diproses oleh sistem:



Gambar 4.1 Diagram Alir Sistem

Pada gambar 4.1 dapat dicermati bahwa sistem membutuhkan data latih dan data uji untuk dapat melakukan proses hingga selesai. Data tersebut dimasukkan oleh *user*, data latih terlebih dahulu disimpan dalam *database*. Data yang disimpan merupakan hasil ekstraksi fitur dengan metode MFCC. Selanjutnya hasil ekstraksi fitur data latih dan data uji dibandingkan dengan menggunakan metode DTW. Sehingga, didapatkan hasil perhitungan jarak kedua data. Sistem akan membandingkan data uji dengan semua data dalam *database,* data yang memiliki hasil perhitungan terkecil merupakan hasil keluaran sistem.

V. HASIL DAN PEMBAHASAN

5.1 Implementasi Antar Muka

1. Menu Utama

Menu utama adalah menu yang tampil pertama kali saat aplikasi dijalankan. Menu ini dibuat berdasarkan rancangan yang telah dibuat sebelumnya.

Gambar 5.1 Menu Utama

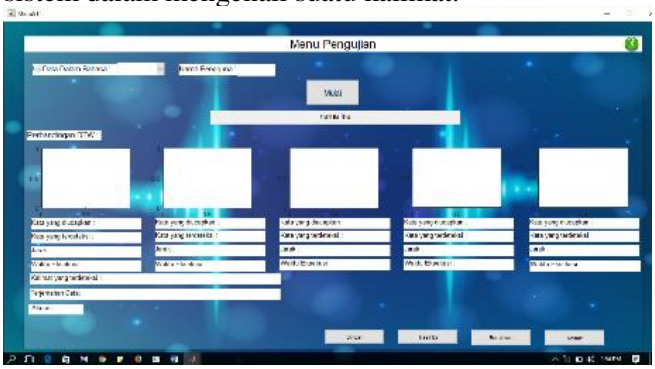
2. Menu Pengujian 1 Kata

Menu pengujian 1 kata akan muncul saat *button* pengujian 1 kata dipilih. Pada menu ini, *user* dapat melakukan pengujian sistem dalam mengenali suatu kata. Gambar 5.2 merupakan tampilan menu pengujian kata.

Gambar 5.2 Menu Pengujian 1 Kata

3.Menu Pengujian 1 Kalimat

Menu pengujian 1 kalimat akan muncul saat *button* pengujian 1 kalimat dipilih. Pada menu ini, *user* dapat melakukan pengujian sistem dalam mengenali suatu kalimat.



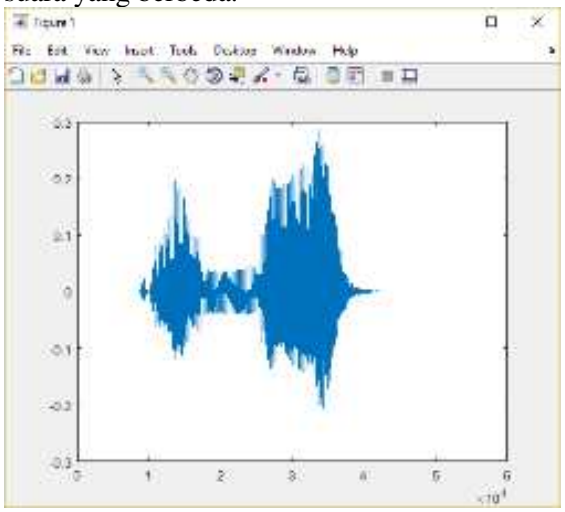
Gambar 5.3 Menu Pengujian 1 Kalimat

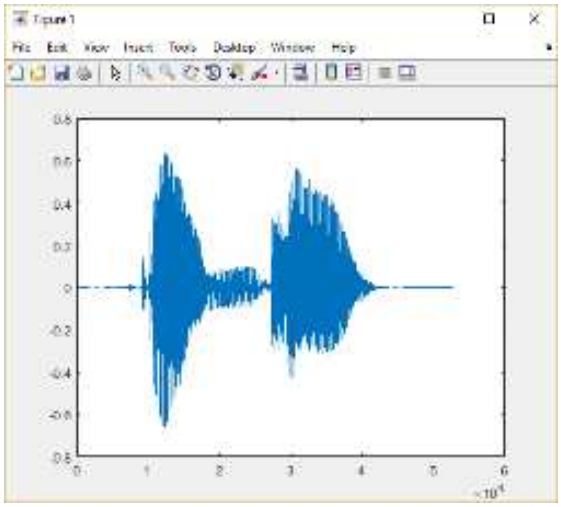
5.2 Pembahasan

Berdasarkan penelitian yang telah dilakukan, penyebab kegagalan pengenalan ucapan dapat dianalisis, pada umumnya disebabkan oleh :

a. Banyaknya variasi sinyal suara suatu kata yang sama

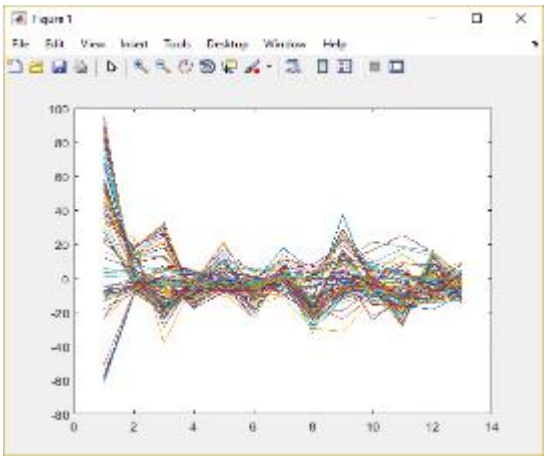
Berdasarkan pengujian yang telah dilakukan, kesensififitas sinyal suara memiliki pengaruh yang besar dalam sistem terutama dalam hal akurasi. Gambar 5.33 merupakan suatu kata yang sama diucapkan oleh orang yang sama, tetapi memiliki hasil pengenalan suara yang berbeda.

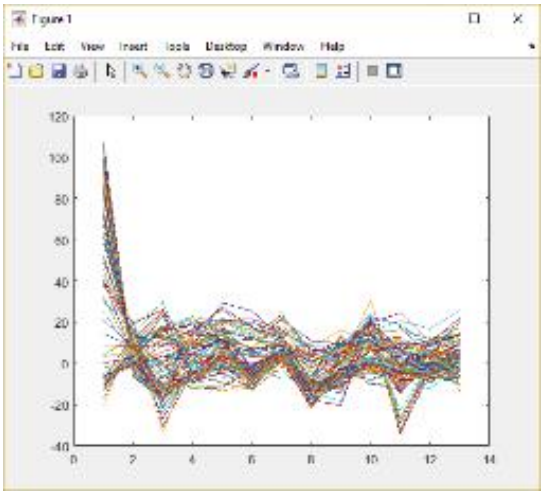




Gambar 5.4 contoh kata ‘anda’ yang memiliki bentuk sinyal yang berbeda

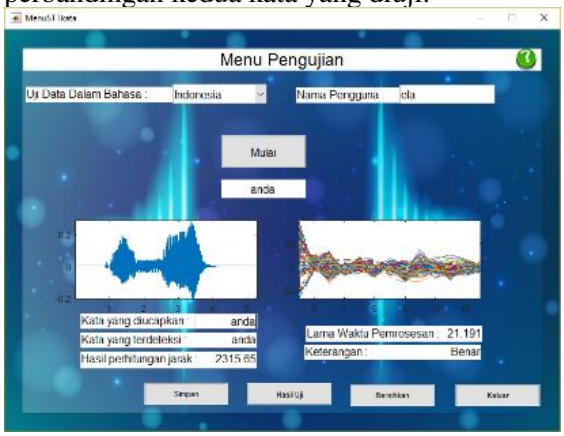
Dapat dilihat pada gambar 5.4, bahwa walaupun orang yang sama mengucapkan kata yang sama, sinyal suara yang terbentuk jauh berbeda. Hal ini dikarenakan setiap orang sulit mengucapkan suatu kata dengan bunyi yang persis sama. Sehingga data masukan dari *user* baik itu pelafalan atau variasi tinggi rendah suara sangat mempengaruhi hasil yang dikeluarkan oleh sistem. Gambar 5.5 merupakan hasil ekstraksi dari kedua sinyal suara.



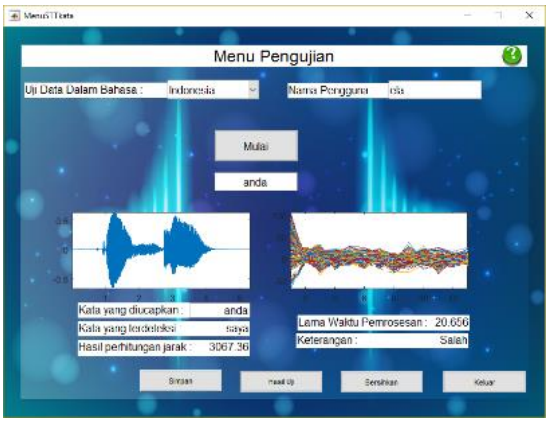


Gambar 5.5 Hasil Ekstraksi Sinyal

Ketika diuji pada sistem kedua kata ‘anda’ tersebut memiliki hasil yang berbeda. Gambar 5.5 dan gambar 5.6 merupakan hasil perbandingan kedua kata yang diuji.



Gambar 5.5 Kata ‘Anda’ yang dikenali

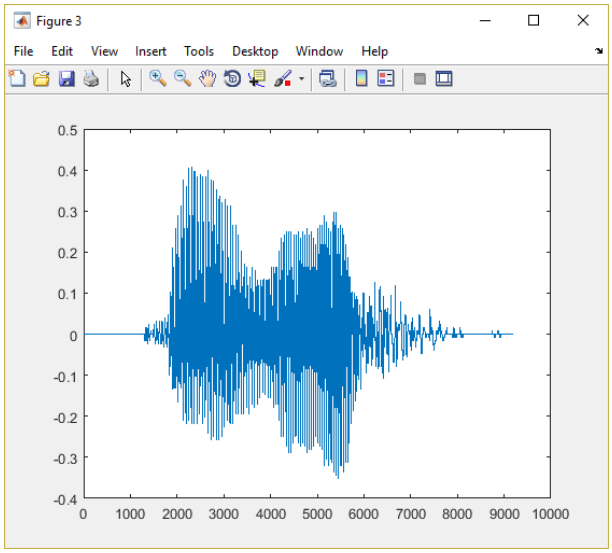


Gambar 5.6 Kata ‘Anda’ yang tidak dikenali

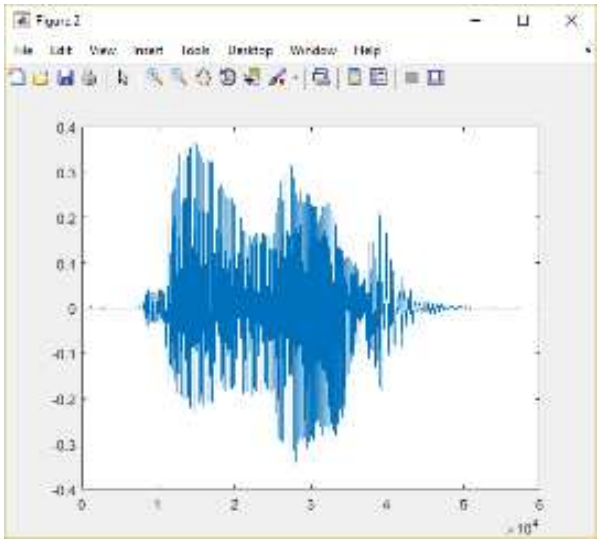
Pada gambar 5.35 kata ‘anda’ yang pertama memiliki jarak 2315.65 terhadap data tersimpan yang dibandingkan, sedangkan kata ‘anda’ yang kedua memiliki jarak 3200.37. Kata ‘anda’ yang kedua pada gambar 5.36 memiliki jarak lebih dekat dengan kata ‘saya’ yaitu 3067.36 sehingga sistem menampilkan hasil pengujian yaitu kata ‘saya’. Hal ini disebabkan karena data kata ‘anda’ yang tersimpan di *database* lebih memiliki kemiripan dengan kata ‘anda’ yang pertama.

b. Kualitas suatu sinyal suara.

Kualitas suatu sinyal suara merupakan salah satu factor yang dapat mempengaruhi pengenalan suatu sinyal tersebut. Hal tersebut dikarenakan semakin kecil ukuran sinyal suara (dalam hal ini frekuensi) maka semakin sedikit data sinyal suara yang dimilikinya. Misal terdapat gambar 4 bit dan gambar 16 bit, gambar 4 bit tentunya akan terlihat lebih buram karena sedikitnya nilai informasi yang terdapat pada gambar tersebut. Berikut merupakan contoh sinyal suara 8000 Hz dan 48000 Hz.



Gambar 5.2 Kata ‘Boleh’ dengan frekuensi 8 kHz



Gambar 5.3 Kata ‘Boleh’ dengan frekensi 48 kHz

Dapat dilihat pada gambar 5.37, sinyal suara yang dihasilkan memiliki frekuensi yang lebih kecil dibanding gambar 5.38, yaitu sekitar 10 kHz, sedangakan gambar 5.38 memiliki frekuensi 60 kHz. Sinyal suara pada gambar 5.37 tidak dapat diuji ke dalam karena memiliki format file .wav, sedangkan sistem yang dibuat memiliki batasan format file yaitu m4a.

Secara umum implementasi metode DTW ke sistem *Speech to Text* dapat dikatakan berhasil karena memiliki tingkat akurasi yang cukup tinggi yaitu sebesar 95.85% terhadap 217 data yang diuji dan tingkat kesalahan sebesar 4,15% untuk pengujian data sinyal suara satu kata. Sedangkan, untuk pengujian data sinyal suara dengan menggunakan satu kalimat yang terdiri dari 5 kata, terhadap 50 data yang diuji, sistem berhasil mengenali dengan baik 47 kalimat dengan rata-rata akurasi sebesar 94%, 3 kalimat lainnya tidak dapat dikenali dengan sempurna atau ia memiliki rata-rata tingkat kesalahan sebesar 6%

VI. PENUTUP

6.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan sebelumnya, maka kesimpulan yang dapat diambil adalah bahwa penelitian ini telah menghasilkan sebuah sistem *speech to text* yang menerapkan algoritma *dynamic time warping*. Berdasarkan hasil pengujian data sinyal suara, untuk pengujian dengan satu kata memiliki jumlah total 217 data yang diuji, sistem mampu mengenali sebanyak 208 kata dengan rata-rata akurasi sebesar 95.85% dan tingkat kesalahan sebesar 4,15 % atau sebanyak 9 kata. Sedangkan, untuk pengujian data sinyal suara dengan menggunakan satu kalimat yang terdiri dari 5 kata, terhadap 50 data yang diuji, sistem berhasil mengenali dengan baik 47 kalimat dengan rata-rata akurasi sebesar 94%, 3 kalimat lainnya tidak dapat dikenali dengan sempurna atau ia memiliki rata-rata tingkat kesalahan sebesar 6%.

6.2 Saran

Berdasarkan hasil dan pembahasan dapat dikatakan sistem ini belum memiliki kesempuranaan yang baik, untuk itu diharapkan adanya saran guna pengembangan penelitian lebih lanjut. Adapun beberapa saran tersebut adalah perlu dibangun sistem yang memiliki data dalam jumlah yang lebih banyak, sehingga sistem ini dapat mengenali kata yang memiliki banyak variasi, misal dari segi intonasi pengucapan ataupun pelafalan yang tidak stabil. Kemudian perlu ditambahkan pengklasifikasian data menggunakan metode *clustering* dalam pencocokan fitur data pada sistem, sehingga sistem tersebut diharapkan mampu untuk mempersingkat waktu pemrosesan data. Perlu ditingkatkan pula kemampuan sistem utuk dapat menguji data dengan format file yang berbeda.

DAFTAR PUSTAKA

[1] S. Paulson and B. Thilagavathi, "An Adaptable Speech to Sign Languange Translation Sistem," *International Journal of Engineering Research & Technology (IJERT),* vol. 3, no. 3, p. 1813, 2014.

[2] S. Swamy and K. Ramakrishnan, "An Efficient Speech Recognition System," *Computer Science & Engineering: An International Journal (CSEIJ),* vol. 3, no. 4, pp. 21-27, August 2013.

[3] X. Han, "Gesture and Voice Control of Internet of Things," Electronics and Computer Engineering at Massey University, Auckland, New Zealand, 2015.

[4] S. D. Dhingra, G. Nijhawa and P. Pandit, "Isolated Speech Recognition Using MFCC and DTW," *International Journal of Advanced Research in Electrical , Electronic and Instrumentation Engineering,* vol. 2, no. 8, pp. 4085-4092, 8 August 2013.

[5] S. W. Smith, "The Scientist and Engineer's Guide to Digital Signal Processing," California Technical Publishing, 1997. [Online]. Available: http://www.dspguide.com/ch1.htm. [Accessed 26 October 2016].

[6] A. Sannino, "Analyzing Discontinuous Speeech in EU Conversation : A Methodological Proposal," *Journal of Pragmatic,* vol. 38, pp. 543-566, 2006.

[7] K. Chakraborty, A. Talele and P. S. Upadhya, "Voice Recognition Using MFCC Algorithm," *International Journal of Innovative Research in Advanced*

Engineering (IJIRAE), pp. 158-161, 2014.

[8] L. Jalan, R. Masram, R. Jadhav and T. Palav, "Speech Recognition Based Learning System," *International Journal of Engineering Trens and Tehcnology,* vol. 4, no. 2, pp. 165-169, 2013.

[9] A. W. B. H. &. T. D. Dennis, Sistem Analysis and Design with UML Version 2.0, United States of America: John Willey & Sons, Inc., 2005.

[10] H. Arman, "Analisa Performance Metode Gabor Filter Untuk Pengenalan Wajah," Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 2012.

[11] N. S. Sukmadinata, Metode Penelitian Pendidikan, Bandung: Rosda, 2005.

[12] Sukardi, Metodologi penelitian pendidikan kompetensi dan praktiknya., Jakarta: PT. Raja Grafindo Persada, 2003.