

Pengenalan Pola Ucapan Kata Menggunakan Metode *Dynamic Time Warping* (DTW) Berbasis Multimedia Interaktif

Youllia Indrawaty N^[1], Rio Korio Utoro^[1], Reza Dio Nugraha^[1]

^[1]Jurusan Teknik Informatika, Fakultas Teknologi Industri
Institut Teknologi Nasional Bandung

youllia@itenas.ac.id, riokorio@itenas.ac.id, rezadionugraha@gmail.com

ABSTRAK

Pengenalan pola suara dapat dimanfaatkan oleh tunarungu sebagai media pembelajaran alternatif. Aplikasi ini berfungsi untuk memberikan penilaian kecocokan pelafalan suara bagi anak tunarungu. Salah satu metode dalam mengenali pola suara adalah *Dynamic time warping* (DTW). DTW merupakan salah satu bentuk pemodelan suara, yang digunakan untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian dengan yang tersedia pada sinyal referensi. Pada aplikasi ini DTW digunakan sebagai pengukur kecocokan suara anak tunarungu dengan suara referensi. Untuk dapat mengenali pola suara dibutuhkan masukan berupa suara yang rekam oleh anak tunarungu, kemudian dilakukan proses ekstraksi ciri suara dengan metode *Linear Predictive Coding* (LPC). Kemudian parameter LPC tersebut digunakan pada proses pencocokan dengan suara referensi yang telah disimpan dengan menggunakan metode DTW. Nilai kecocokan digunakan sebagai nilai evaluasi pelatihan pengucapan kata. Model skenario multimedia merupakan alur komponen dari awal multimedia dimulai hingga multimedia selesai. Sehingga dapat menghasilkan suatu tampilan yang interaktif. Model firefly merupakan salah satu model skenario multimedia yang bentuknya menyerupai kunang-kunang. Perancangan aplikasi ini menggunakan model skenario Firefly, karena penyajian informasi sesuai dengan alur kejadian yang ditampilkan. Tingkat keberhasilan aplikasi dalam mengenali pola terhadap suara ucapan anak tunarungu sebesar 77,5%.

Kata kunci : Tunarungu, Pengenalan pola, *Linear Predictive Coding*, , *Dynamic Time Warping*, Multimedia, firefly

ABSTRACT

Pattern voice recognition can be used by deaf children as alternative learning media. The Function of application are give a similarity rate pronunciation for deaf children. One of the method pattern voice recognition is *Dynamic time warping* (DTW). DTW used to accommodate differences in timing between record process with available references signal. In this application DTW used to measure voice deaf children with references voice. For identify pattern voice recognition require record voice of deaf children, and then feature extraction process voice with *Linear Predictive Coding* (LPC). And then LPC parameters used in similiarity process with stored references voice using DTW. Similiarity rate as prononciation word learning evaluation. Multimedia scenario model is component step from start until end multimedia. Firefly model is one of the multimedia scenarios which phase like a firefly. This application design is using firefly scenario, because presentation of information is suitable with step of event displayed. The success rate of the application in recognizing the pattern of the words sound of deaf children is 77,5%.

Keyword: deaf, Pattern recognition, *Linear Predictive Coding* (LPC), *Dynamic Time Warping* (DTW), Multimedia, firefly

PENDAHULUAN

Secara umum tunarungu dapat diartikan seseorang yang mengalami hambatan dalam mendengar, banyak pengertian yang dikemukakan oleh para ahli tentang pengertian tunarungu salah satunya yaitu menurut Somantri (2007:93) tunarungu dapat diartikan sebagai suatu keadaan kehilangan pendengaran yang mengakibatkan seseorang tidak dapat menangkap berbagai rangsangan, terutama melalui indera pendengarannya.

Anak tunarungu kecenderungan memiliki permasalahan dalam belajar yang disebabkan oleh kurangnya kemampuan untuk memahami ujaran orang lain. Kondisi ini, sering disebabkan oleh ketidakmampuan dalam penguasaan pendengaran, sehingga akhirnya mereka menghindari aktivitas berbahasa karena pendengarannya terganggu.

Bina Komunikasi Persepsi Bunyi dan Irama atau yang disingkat BKPBI merupakan pembinaan dalam penghayatan bunyi yang dilakukan dengan sengaja atau tidak sengaja, sehingga sisa-sisa pendengaran dan perasaan vibrasi yang dimiliki anak-anak tunarungu dapat dipergunakan secara optimal untuk berinteraksi dengan sekitarnya.

Komunikasi merupakan salah satu materi dalam pelajaran BKPBI. Materi komunikasi berisi latihan cara mengucapkan kata-kata, diharapkan melalui materi komunikasi anak tunarungu dapat mengucapkan kata-kata dengan jelas, sehingga dapat dimengerti oleh lawan bicara yang bukan tunarungu.

Salah satu bentuk pengenalan pola suara adalah *Dynamic time warping* (DTW). DTW sering digunakan dalam area: *voice recognition, handwriting and online signature matching, sign language recognition and gestures recognition, data mining and time series clustering, computer vision and computer animation, surveillance, protein sequence alignment and chemical engineering, dan music and signal processing.*

Disimpulkan bahwa dengan pelajaran BKPBI diharapkan dapat menjadi sebuah alternatif melatih kemampuan berbicara anak tunarungu, sehingga anak tunarungu dapat merangsang sisa pendengaran. Pemanfaatan teknologi sebagai media pembelajaran dapat dilakukan untuk membuat sebuah alternatif dalam proses belajar mengajar pelajaran BKPBI, melalui aplikasi multimedia interaktif dengan materi komunikasi pada pelajaran BKPBI yang berisi latihan mengucapkan kata dengan cara pengenalan pola menggunakan algoritma DTW. Oleh karena itu dibuatlah sebuah alternatif pembelajaran BKPBI dengan membangun aplikasi multimedia interaktif latihan mengucapkan kata untuk anak tunarungu.

Rumusan Masalah

Adapun masalah yang dirumuskan adalah sebagai berikut :

1. Bagaimana cara membangun aplikasi multimedia interaktif mengucapkan kata pada pelajaran BKPBI.
2. Bagaimana cara implementasi algoritma DTW pada proses pengenalan pola suara aplikasi multimedia interaktif mengucapkan kata.

Tujuan

Tujuan dari penelitian ini adalah untuk menerapkan algoritma DTW pada pengenalan pola ucapan bagi anak tunarungu

Batasan Masalah

Dalam penelitian ini batasan masalah adalah sebagai berikut:

1. Menggunakan *Headphone* dengan volume suara diatas 80 dB untuk mendengarkan suara pengucapan kata.
2. Pengguna merupakan anak tunarungu yang memiliki tingkat tunarungu ringan(intensitas 20-45 dB) hingga tingkat tunarungu sedang(intensitas 46-70 dB).
3. Latihan pengucapan kata menggunakan kata bahasa Indonesia,

hanya menggunakan kata terisolasi (*isolated word*), dan hanya menggunakan 20 contoh kata.

LANDASAN TEORI

Pengenalan Pola^[1]

Pengenalan pola (*pattern recognition*) dapat diartikan sebagai proses klasifikasi dari objek atau pola menjadi beberapa kategori atau kelas. Dan bertujuan untuk pengambilan keputusan (Theodoridis and Koutroumbas 2006, 1).

Pola adalah bentuk atau model (atau, lebih abstrak, suatu set peraturan) yang dapat dipakai untuk membuat atau untuk menghasilkan suatu atau bagian dari sesuatu, khususnya jika sesuatu yang ditimbulkan mempunyai sejenis pola dasar yang dapat ditunjukkan atau terlihat, yang mana dapat dikatakan mempertunjukkan pola.

Deteksi pola dasar disebut pengenalan pola. Untuk dapat mengenali suara dan pemilik suara (*speaker*), diperlukan suatu pengenalan pola yang dapat mengenal dan mencocokkan pola masukan dan pola yang telah disimpan di dalam *database*.

Pengenalan pola merupakan bidang dalam pembelajaran mesin dan dapat diartikan sebagai tindakan mengambil data mentah dan bertindak berdasarkan klasifikasi data. Dengan demikian, hal tersebut merupakan himpunan kaidah bagi pembelajaran yang diawasi (*supervised learning*).

Dynamic Time Warping (DTW)^[4]

Dynamic time warping (DTW) adalah algoritma yang menghitung *warping path* yang optimal antara dua *sekuens numerik* data sehingga *output* adalah nilai-nilai *warping path* dan jarak diantara kedua *sekuens numerik* tersebut.

Algoritma DTW disebut juga sebagai *non-linear sequence alignment*, sehingga algoritma ini lebih realistis untuk digunakan dalam mengukur kemiripan suatu pola (*pattern/template matching*).

Misalnya ada dua buah *sekuens numerik* ($x_1, x_2, x_3, \dots, x_n$) dan ($y_1, y_2, y_3, \dots, y_n$) saja panjang *sekuens numerik* data bisa berbeda, berarti $n=m$ atau $n \neq m$. Tahap awal dari algoritma DTW adalah menghitung jarak local (*local distance*) antar elemen dari kedua *sekuens numerik* menggunakan teknik penghitungan jarak yang berbeda, biasanya yang digunakan adalah Euclidean Distance. *Local distance* dihitung dengan mencari nilai absolute dari selisih kedua *sekuens numerik* data yang dirumuskan dengan :

$$d_{ij} = |x_i - y_j|, i = 1:n \text{ dan } j = 1:m \dots 1$$

Jika dalam matriks maka dapat ditulis dengan memiliki n garis dan m kolom.

$$x_{ij} = d_{ij} + \min(x_{i-1,j-1}, x_{i-1,j}, x_{i,j-1}) \dots 2$$

Dimana x merupakan jarak minimal antara subsekuens. Langkah selanjutnya merupakan *point* yang harus kita lewati untuk menemukan *warping path*. *Warping path* adalah sebuah *path* yang melewati jarak matrik minimum dari elemen x_{ij} ke x_{11} . Ongkos *warping path* secara global dari dua *sekuens*:

Dimana w_i adalah elemen yang dimiliki *warping path* dan p adalah jumlahnya. Penghitungannya dibuat untuk dua *sekuens*.

$$GC = \frac{1}{p} \sum_{i=1}^p w_i \dots 3$$

Linear Predictive Coding (LPC)^[2]

Linear Predictive Coding atau LPC adalah sebuah alat yang digunakan dalam sebagian besar pemrosesan sinyal audio dan pemrosesan wicara untuk mewakili *spectral envelope* dari sinyal digital pembicaraan dikompresi, menggunakan informasi dari *linear input model*. LPC adalah salah satu yang paling baik dalam teknik analisi wicara, dan salah satu metode yang paling berguna untuk *encoding* berkualitas pada *bit-rate* yang rendah dan menyediakan yang baik mengenai perkiraan pembicaraan parameter LPC juga termasuk cukup efisien untuk

ANALISIS DAN PERANCANGAN

Analisis Kebutuhan User

Anak tunarungu adalah anak yang mengalami gangguan pada organ pendengarannya sehingga mengakibatkan ketidakmampuan mendengar, mulai dari tingkatan yang ringan sampai yang berat sekali yang diklasifikasikan kedalam tuli dan kurang dengar.

Tunarungu berdasarkan kehilangan daya dengarnya dibagi menjadi:

- a. Kurang Dengar (< 90 dB)
 - Ringan : intensitas 20-45 dB.
 - Sedang : intensitas 46-70 dB.
 - Berat : intensitas 71-90 dB.
- b. Tuli (> 90 dB)

Aplikasi ini dirancang bagi anak tunarungu yang memiliki tingkat kehilangan daya dengar dengan tingkat ringan hingga tingkat sedang, karena aplikasi ini mempunyai fitur agar pengguna bisa mendengarkan contoh pengucapan kata, sehingga diharapkan suara pengucapan bisa terdengar maksimal agar pengguna dapat mengucapkan kata dapat sesuai contoh ucapan kata.

Analisis Kebutuhan Sistem

Pada tahap ini dilakukan analisa terhadap fitur yang akan diterapkan didalam pembangunan aplikasi. Proses analisa dilakukan dengan menganalisa cara kerja pengenalan pola (*Pattern Recognition*) suara ucapan kata. Secara garis besar hasil dari analisa terhadap sistem dapat dilihat pada blok diagram Gambar 3.

A/D converter

Proses konversi *Analog to Digital* telah dilakukan oleh *soundcard*.

Preemphasis

Preemphasis adalah suatu proses perbaikan sinyal suara ucapan dari gangguan *noise*, sehingga dapat meningkatkan tingkat akurasi pengenalan suara untuk tahap selanjutnya, default dari nilai α yang digunakan dalam proses *preemphasis* adalah 0.95. Rumus yang digunakan untuk proses *preemphasis* :

$$y(n) = s(n) - \alpha s(n-1) \dots\dots\dots 7$$

Contoh perhitungan *preemphasis*:

Sinyal : [100, 150, 200]

$s_1=100, s_2=150, s_3=200, \alpha=0,95$

$Y(1)=100 - (0,95*0) = 100$

$Y(2)=150 - (0,95*100) = 55$

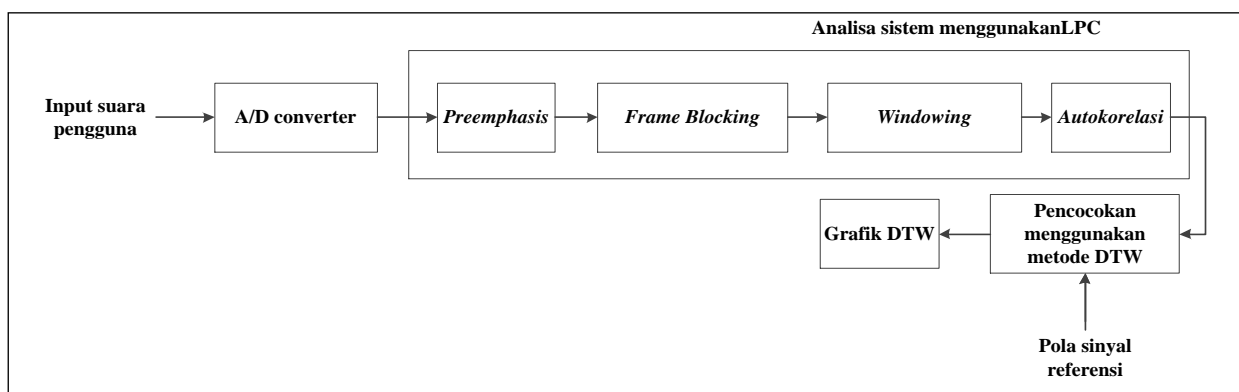
$Y(3)=200 - (0,95*150) = 57,5$

Hasil sinyal setelah *preemphasis* adalah [100, 55, 57,5]

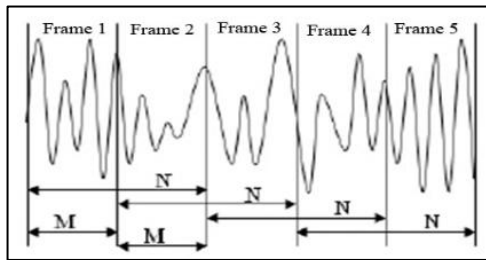
Frame Blocking

Pembingkaiian sinyal (*Framing*) adalah suatu proses setelah melakukan *preemphasis*, sinyal suara yang telah di *preemphasis* dibagi menjadi beberapa *frame* agar memudahkan dalam perhitungan dan analisa suara, setiap potongan disebut sebagai *frame*.

Satu *frame* terdiri dari beberapa sampel. Hasil *frame blocking* seperti pada gambar 6, dimana M merupakan panjang tiap *frame* dan N adalah *overlapping* tiap *frame*.



Gambar 3 Blok Diagram Sistem Secara Umum



Gambar 4 Contoh Hasil Frame

Blocking

Contoh perhitungan penjendelaan sinyal

$$s = 10\text{ms} = 0,01\text{s}$$

$$f_s = 8000\text{Hz}$$

$$N = f_s * s = 8000 * 0,01 = 80$$

$$\text{overlapping frame (M)} = 80$$

$$x(n) = y(M+n) \dots \dots \dots 8$$

$$x(1) = 100(80+1) = 8100$$

Windowing

Windowing bertujuan untuk mengurangi terjadinya *aliasing* yang mana merupakan suatu efek dari timbulnya sinyal baru yang memiliki frekuensi yang berbeda dengan sinyal aslinya. Efek tersebut dapat terjadi karena rendahnya jumlah *sampling rate*. Hal ini dilakukan pada bagian awal dan akhir sinyal. Jika di definisikan sebuah *window* $w(n)$ dan sinyal tiap bagian adalah i . Contoh sinyal hasil *windowing* seperti pada gambar 7.

Jenis *window* yang digunakan adalah *Hamming Window* dengan rumus sebagai berikut:

$$W[n] = 0,54 - 0,46 * \cos(2 * \pi * n / N - 1) \dots \dots \dots 9$$

Contoh perhitungan *Hamming window* :

$$n = 1, 2, 3, \dots$$

$$\pi = 3,14$$

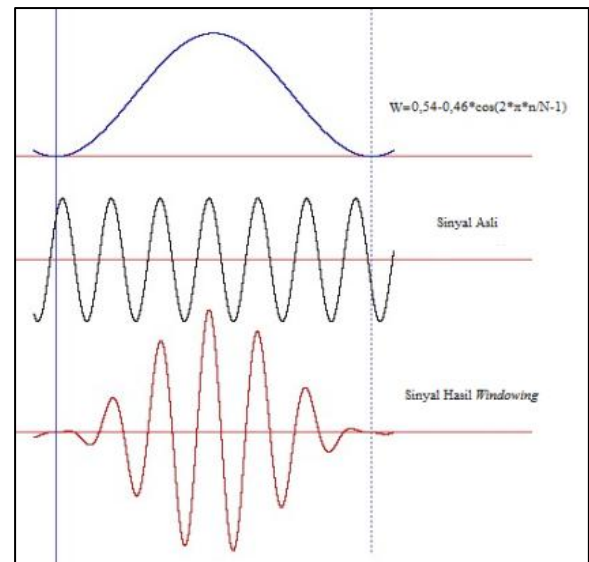
$$\text{Panjang sinyal (N)} = 8000$$

$$w(n) = 0,54 - 0,46 \cos(2 * 3,14 / (N-1))$$

$$w(1) = 0,54 - 0,46 \cos(8)$$

$$w(1) = 0,54 - 0,46 * 1$$

$$w(1) = 0,08 \text{ dst (untuk } n = 1, 2, 3, \dots)$$



Gambar 5 Contoh Hasil Windowing

Autokorelasi

Pada proses ini masing-masing *frame* yang telah melalui proses *windowing* kemudian diproses autokorelasi dengan nilai *autokorelasi* yang tertinggi adalah orde dari analisa LPC, biasanya orde LPC tersebut 8 sampai 24.

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n) \tilde{x}_l(n+m)$$

Metode Dynamic Time Warping (DTW)

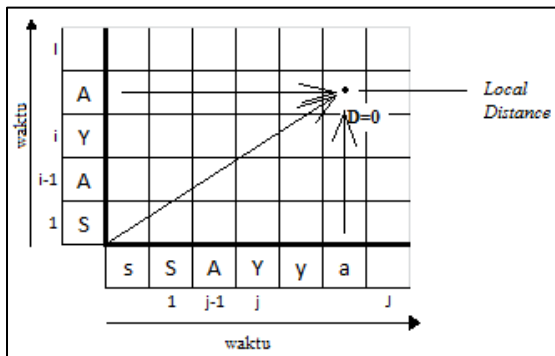
Metode DTW merupakan proses akhir pengenalan pola ucapan kata diproses. DTW merupakan metode untuk mencocokkan pola ucapan kata pengguna, dengan pola ucapan sinyal referensi dalam menentukan kesamaan jarak antara pola-pola yang berbeda dan menghitung nilai *warping path* dari dua waktu dan jaraknya.

DTW terbagi menjadi tiga proses utama yaitu *local distance*, *accumulated distance*, dan *the distance*.

1. *Local Distances* :

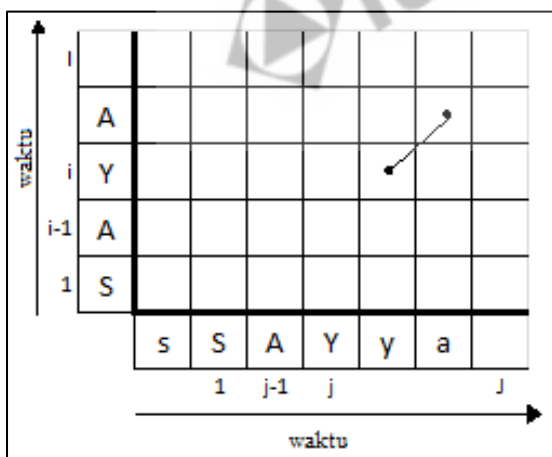
Local distances adalah jumlah jarak *Euclidean*, *Euclidean* adalah digunakan untuk mengukur jarak (*distance*). *Distance* adalah suatu ukuran yang dipakai untuk menentukan kemiripan dua buah data. Pertama-tama *local distance* menghitung biaya minimum dan jalur dari

kata ucapan. *Local distance* dapat dilihat seperti pada gambar 8



Gambar 6 Diagram *Local Distance*
2. *Accumulated Distance*

Setelah mendapatkan jalur optimal pola suara, panjang *frame* dan nilai *spectrum* dari tiap *frame* selanjutnya adalah proses *Accumulated Distance*. *Accumulated Distance* merupakan perhitungan jarak antara *local distance* satu dengan *local distance* yang lainnya. *Accumulated distance* dapat dicari dengan mencari jarak minimal antara *local distance* ditambah *accumulated distance* disampingnya. *Accumulated distance* dapat dilihat seperti pada gambar 9.

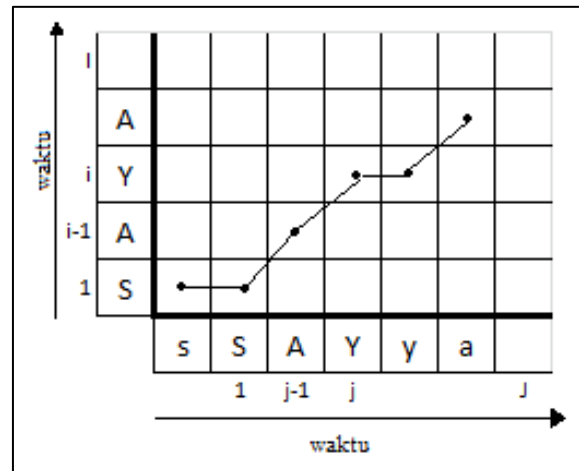


Gambar 7 Contoh Hasil Proses
Accumulated Distance

3. *The Distances*

The Distance adalah jarak akhir yang menunjukkan *warping path* jarak antara dua suara masukan dengan jarak suara

referensi. Hasil *the distance* berupa *warping path* seperti pada gambar 10.



Gambar 8 Hasil Proses *The Distance*

Storyline


Sistem yang dibuat berupa multimedia interaktif latihan pengucapan kata merupakan salah satu media alternatif pembelajaran bagi anak tunarungu dalam melatih kemampuan mengucapkan kata. Pada menu utama pengguna memilih jenis kata, kemudian akan masuk menu pilihan kata yang termasuk dalam jenis kata yang dipilih sebelumnya. Pengguna memilih salah satu kata yang akan dilatih kemudian masuk ke menu latihan, akan terlihat grafik sinyal suara kata, sedangkan ketika pengguna menekan tombol mulai, sistem akan otomatis merekam suara pengguna dalam dua detik. Grafik sinyal suara ucapan pengguna akan terlihat, untuk melihat hasil proses pengenalan pola kata, dengan menekan tombol hasil. Jika suara ucapan pengguna sesuai maka akan muncul poin dan bisa melihat grafik *spectrogram* dan grafik *DTW*, jika tidak sistem akan meminta untuk mengulang kembali proses perekaman suara. Pengguna akan mendapatkan tanda ceklis pada kata yang telah sesuai pada halaman daftar kata.

Storyboard

Storyboard adalah penyelenggara grafis seperti serangkaian ilustrasi atau

gambar yang ditampilkan secara berurutan untuk tujuan pra-memvisualisasikan sebuah gambar bergerak, animasi, gerak grafis atau urutan media interaktif. Contoh *storyboard* seperti pada tabel 1

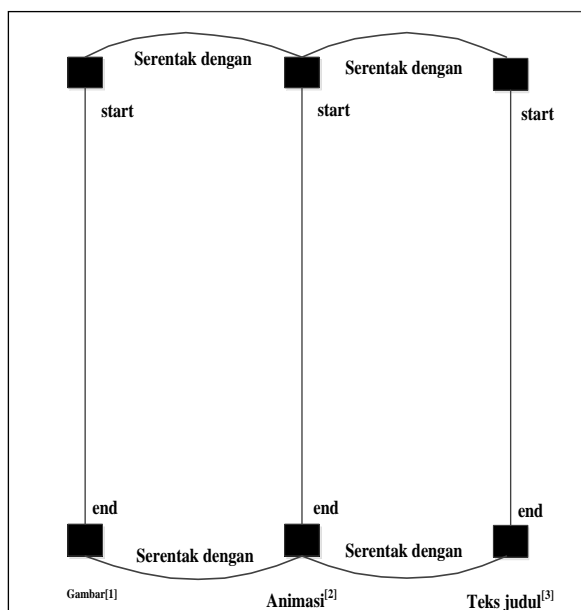
Tabel 1 *Storyboard*

NO	Ilustrasi	Keterangan
1		Tampilan halaman judul, terdiri dari: - Judul aplikasi - Animasi judul

Model Skenario Multimedia Interaktif Menggunakan Firefly

Model skenario yang digunakan pada pembangunan aplikasi ini adalah model *firefly*. Model *firefly* merupakan model skenario multimedia yang bentuknya menyerupai kunang-kunang.

Berdasarkan *storyboard*, maka dapat dibuat model skenario *firefly* gambar 11 merupakan salah satu contoh penggunaan skenario *firefly*

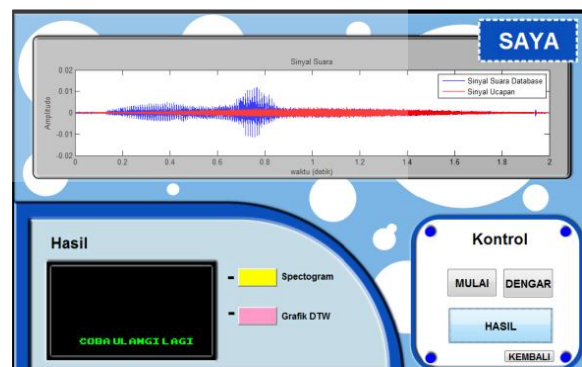


Gambar 9 *Firefly* Halaman Judul

IMPLEMENTASI DAN PENGUJIAN

Halaman Latihan Pengucapan

Halaman latihan pengucapan seperti Gambar 12 menampilkan lima tombol yaitu tombol mulai, dengar, hasil, spektrogram, grafik DTW kembali dan gambar sinyal suara referensi. Tombol mulai berfungsi untuk merekam suara pengguna selama dua detik, Tombol dengar berfungsi untuk mendengarkan suara pengucapan kata, Tombol hasil berfungsi untuk melihat seberapa besar kecocokan antara sinyal suara pengguna dengan sinyal suara referensi, Tombol spektrogram berfungsi untuk melihat grafik spektrogram, Tombol grafik DTW berfungsi untuk melihat hasil dari metode DTW yang berupa warping path antara sinyal suara pengguna dengan sinyal referensi, sedangkan tombol kembali berfungsi untuk kembali ke halaman jenis kata.



Gambar 10 Halaman Latihan Pengucapan

Pengujian *Beta* ini dilakukan dengan menguji interaksi aplikasi dan pengguna. Jumlah pengguna dalam pengujian ini adalah 10 orang siswa SMP LB dan lima orang bukan tunarungu. Adapun hasil pengujian yang ditampilkan hanya tiga

orang anak tunarungu dan tiga orang bukan tunarungu adalah sebagai berikut.

Tabel 2 Pengujian interaksi

No	Nama	Kata	Poin Kesesuaian	Suara terdengar?
1	Bryan	makan	55	K
		baju	-	K
		saya	45	TJ
		sakit	33	TJ
2	Chika	makan	49	TJ
		baju	42	TJ
		saya	50	TJ
		sakit	41	TJ
3	Zasqia	makan	36	TJ
		baju	44	TJ
		saya	58	K
		sakit	-	TJ

aplikasi dan pengguna tunarungu

Keterangan

SJ =SuaraTerdengar Sangat Jelas

J = Suara Terdengar Jelas

K = Suara Terdengar Kecil

TJ=Suara Terdengar Tidak Jelas

Tabel 3 Pengujian interaksi aplikasi dan pengguna bukan tunarungu.

No	Nama	Kata	Poin Kesesuaian	Suara terdengar?
1	Andhika	makan	36	J
		baju	44	J
		saya	41	J
		sakit		J
2	Zidni	makan	42	J
		baju	40	J
		saya	33	J
		sakit	36	J
3	Gen	makan	31	J
		baju	43	J
		saya	30	J
		sakit	31	J

Keterangan

SJ =SuaraTerdengar Sangat Jelas

J = Suara Terdengar Jelas

K = Suara Terdengar Kecil

TJ=Suara Terdengar Tidak Jelas

Berdasarkan tabel 2Tingkat keberhasilan aplikasi terhadap suara anak tunarungu adalah 31 kali dari 40 kali pengujian, dengan persentase sebesar $(100/40)*31= 77,5 \%$

Rata-rata poin kesesuaian :

Rata-rata = (jumlah poin) / jumlah pengujian kata

jumlah pengujian tiap kata=10

- Kata 'makan'
 $= (55+49+36+44+45+0+44+45+0+44)/10 = 364/10 = 36,2$
- Kata 'baju'
 $= (0+42+44+51+43+37+39+39+0+39) = 334/10 = 33,4$
- Kata 'saya'
 $= (45+50+58+32+51+45+52+42+32+30) = 437/10 = 43,7$
- Kata 'sakit'
 $= (33+41+0+35+0+31+38+0+0) = 178/10 = 17,8$

KESIMPULAN

Dari hasil implementasi dan pengujian sistem yang telah dilakukan dan ditampilkan pada bab empat, maka diperoleh beberapa kesimpulan, yaitu :

- Aplikasi multimedia interaktif mengucapkan kata telah diuji kepada pengguna dengan reponden 10 orang anak tunarungu dan lima orang pengguna bukan tunarungu.
- Semua fungsi yang diterapkan pada multimedia interaktif latihan mengucapkan kata dengan penerapan metode *Dynamic Time Warping* dan

menggunakan model skenario multimedia *firefly* telah berjalan sesuai dengan yang diharapkan.

DAFTAR PUSTAKA

1. **Theodoridis and Koutroumbas**, 2006, “*Pattern Recognition (Third Edition)*”,
2. **Rabiner, Lawrence and Biing Hwang Juang**, 2006, “*Fundamentals of Speech Recognition*”.
3. **Indrawaty, Youllia**. 2011. *Materi Kuliah Teknik Multimedia :Teknik Multimedia*. Institut Teknologi Nasional : Bandung (tidak dipublikasi).
4. **Sakoe, Hiroaki and Seibi Chiba**. "Dynamic Programming Algorithm Optimization for Spoken Word Recognition" .
5. **Nael Hirzalla, Ben Falchuk, and Ahmed Karmouch**, 1995, *A Temporal Model for Interactive Multimedia Scenarios*”.

