

Rancang Bangun *Text to Sound* menggunakan Metode Hybrid sebagai Sarana membantu Komunikasi Tuna Netra dan Tuna Wicara

Lilik Anifah,* Hainur Rasid Achmadi, dan Dwikoranto
Jurusan Fisika FMIPA Universitas Negeri Surabaya (UNESA)
Jl. Ketintang, Surabaya 60231

Haryanto
Jurusan Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo
Kampus Universitas Trunojoyo, Bangkalan, Madura

Intisari

Alat bantu bagi tuna netra sampai saat ini adalah Reglet dan stylus, Mesin tik braille, Abacus, Peta timbul, Papan baca permulaan, Kertas braille, Tongkat tuna netra, Biji berhitung, Busur derajat braille, Jam dinding bicara, dan Weker talking. Tetapi untuk alat bantu yang bersifat komunikatif yang dapat mengubah informasi teks menjadi informasi suara belum ada. Tujuan Penelitian ini adalah untuk menciptakan suatu perangkat yang dapat mengubah informasi teks (tulisan) menjadi informasi suara dalam bahasa Indonesia. Teks yang akan dibaca diletakkan pada scanner yang selanjutnya diproses sehingga informasi teks tersebut akan otomatis diubah menjadi informasi suara. Hasil dari penelitian ini ternyata sistem yang dibangun telah dapat mengubah informasi yang berupa image (gambar teks) yang diinputkan dari scanner menjadi suara. Masih terdapat keterbatasan sistem ini diharapkan diadakan penelitian lebih lanjut sehingga didapat sistem yang sempurna. Dengan teknologi ini maka diharapkan seseorang yang buta aksara dan tuna netra dapat terbantu dalam membaca suatu teks.

KATA KUNCI: segmentasi teks, normalisasi, teks to speech, sound syntesis

I. PENDAHULUAN

Sejumlah badan dunia seperti UNESCO, UNICEF, WHO, World Bank dan Human Right Watch sangat prihatin dengan kondisi penduduk dunia yang 861 juta diantaranya masih mengalami buta huruf (buta aksara). Ironisnya, 15,04 juta diantaranya berada di Indonesia [1]. Masalah buta huruf sangat terkait dengan kemiskinan, keterbelakangan, kebodohan dan ketidakberdayaan masyarakat. Atas dasar inilah badan-badan internasional gencar mengkampanyekan dan mensosialisasikan pentingnya pemberantasan buta aksara di dunia khususnya negara belahan dunia ketiga seperti Indonesia.

Menurut Dirjen Pendidikan Luar Sekolah, Dr Ace Suryadi kepada Pembaruan di sela acara rangkaian Hari Aksara Internasional di Jakarta, Jumat (24/11) bahwa Indonesia terlibat dalam forum Dakar - Senegal, sebuah forum negara belahan dunia ketiga sepakat dengan point penting akan pengentasan masalah buta aksara ini yakni mengurangi buta aksara sebesar 50 persen pada tingkat orang dewasa pada tahun 2015. Sementara tekad pemerintah sekarang adalah mengurangi hingga tinggal 5 persen pada tahun 2009 dari angka buta aksara yang saat ini di Indonesia masih terdapat 15,04 juta.

Menurut Menteri Pendidikan Nasional (Mendiknas) Bambang Sudibyo, di Jakarta, Rabu (19/9) bahwa jumlah pen-

uduk buta huruf di Indonesia hingga pertengahan tahun 2007 menunjukkan peluang sangat besar untuk mencapai target lima persen pada tahun 2009. Target pemerintah angka buta huruf di Indonesia bisa diturunkan hingga menjadi lima persen pada akhir masa dinas KIB, yaitu tahun 2009 [2].

Selain jumlah buta aksara yang tergolong besar jumlahnya, keterbatasan indera penglihatan tidak jadi halangan bagi penyandang tuna netra untuk maju. Mereka berkemauan besar untuk memperbaiki hidup dengan mengandalkan teknologi informasi (TI). Salut!. Sebagai manusia yang punya panca indera lengkap, kita wajib bersyukur dan berintrospeksi. Tuna netra yang punya kekurangan dalam melihat, punya usaha dan kemauan yang besar. Apalagi usaha mereka untuk memperbaiki kehidupan dengan mengandalkan teknologi informasi. Irwin Dwikustanto, Pimpinan Proyek dari Yayasan Mitra Netra (YMN) mengungkapkan, tuna netra termasuk di dalam kelompok masyarakat yang haknya terdiskriminasi oleh sistem kekuasaan negara dan budaya akibat ketidakseimbangan akses teknologi, komunikasi dan informasi (ICT).

Menurutnya, salah satu ketidakseimbangan itu terletak di bidang teknologi komputer yang merupakan penunjang komunikasi untuk mempermudah pekerjaan. Oleh sebab itu, pihak YMN lalu mengembangkan program strategis untuk tuna netra. Hal itu dilakukan karena program dan disain teknologi informasi (TI) yang ada saat ini, belum cukup memadai untuk komunitas mereka. Pernyataan itu diungkap Aria Indrati, Public Relation YMN. "Dengan akses TI, tuna netra bisa mempunyai peluang yang lebih besar dalam mengerjakan sesuatu.

*E-MAIL: lilikanifah@yahoo.com

Hal itu tentunya berpengaruh besar pada masa depan penyandang kekurangan fisik, seperti masalah pekerjaan,” tuturnya kepada detikinet Rabu(3/8/2005). ”Tuna netra di Indonesia masih terabaikan, padahal mereka juga punya hak yang sama,” kata Aria [3].

Alat bantu bagi tuna netra sampai saat ini adalah Reglet dan stylus, Mesin tik braille, Abacus, Peta timbul, Papan baca permulaan, Kertas braille, Tongkat tuna netra, Biji berhitung, Busur derajat braille, Jam dinding bicara, dan Weker talking. Tetapi untuk alat bantu yang bersifat komunikatif yang dapat mengubah informasi teks menjadi informasi suara belum ada [4].

Pada artikel ini dilaporkan pembuatan suatu perangkat yang dapat mengubah informasi teks (tulisan) menjadi informasi suara dalam bahasa Indonesia. Teks yang akan dibaca diletakkan pada scanner yang selanjutnya diproses sehingga informasi teks tersebut akan otomatis diubah menjadi informasi suara. Dengan teknologi ini maka diharapkan seseorang yang buta aksara dan tuna netra dapat terbantu dalam membaca suatu teks.

II. STUDI PUSTAKA

Penelitian sebelumnya tentang *text to sound* dalam Bahasa Indonesia sangat jarang [5, 6]. Menurut Arry Akhmad Arman pembentukan ucapan itu dimulai dengan adanya hembusan udara yang dihasilkan oleh paru-paru. Cara kerjanya mirip piston atau pompa yang ditekan untuk menghasilkan tekanan udara. Pada saat *vocal cord* berada dalam keadaan tegang, aliran udara akan menyebabkan terjadinya vibrasi pada *vocal cord* dan menghasilkan bunyi ucapan yang disebut *voiced speech sound*. Pada saat *vocal cord* berada dalam keadaan lemas, aliran udara akan melalui daerah yang sempit pada *vocal tract* dan menyebabkan terjadinya turbulensi, sehingga menghasilkan suara yang dikenal sebagai *unvoiced sound*. Cara mengklasifikasikan bagian-bagian atau komponen sinyal ucapan secara sederhana dibagi menjadi tiga keadaan yang berbeda, yaitu

1. Silence, keadaan pada saat tidak ada ucapan yang diucapkan.
2. Unvoiced, keadaan pada saat *vocal cord* tidak melakukan vibrasi, sehingga suara yang dihasilkan bersifat tidak periodik atau bersifat random;
3. Voiced, keadaan pada saat terjadinya vibrasi pada *vocal cord*, sehingga menghasilkan suara yang bersifat kuasi periodik.

Untuk konversi dari teks ke ucapan (*Text to Speech*) pada prinsipnya terdiri dari dua subsistem, yaitu:

1. Bagian konverter teks ke fonem (*Text to Phoneme*) yang terdiri dari *Text Normalization*, *Exception Dictionary* dan *Lookup Letter-to-Phoneme Conversion*, *Prosody Generation*, kemudian *Phonetic Analysis*. Bagian Konverter Teks ke Fonem berfungsi untuk mengubah kalimat masukan dalam suatu bahasa tertentu yang berben-

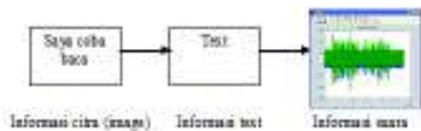
tuk teks menjadi rangkaian kode-kode bunyi yang biasanya direpresentasikan dengan kode fonem, durasi serta pitch-nya. Bagian ini sangat tergantung pada karakter bahasa tertentu. Untuk suatu bahasa yang berbeda bagian ini harus dikembangkan secara lengkap khusus untuk bahasa tersebut. Untuk kondisi yang tidak dapat ditemukan keteraturannya seperti simbol huruf e yang dapat diucapkan sebagai e pepet atau e taling konversinya ditangani oleh bagian *Exception Dictionary Lookup*. Tetapi untuk kondisi yang masih dapat ditangani oleh aturan diimplementasikan oleh bagian *Letter to Phoneme Conversion*.

2. Bagian konverter fonem ke ucapan (*Phoneme to speech*), yaitu *Speech Parameters Generation* dan *Speech Waveform Production*. Bagian Konverter Fonem ke Ucapan akan menerima masukan berupa kode-kode fonem serta pitch dan durasi yang dihasilkan oleh bagian sebelumnya. Berdasarkan kode-kode tersebut, bagian Konverter Fonem ke Ucapan akan menghasilkan bunyi atau sinyal ucapan yang sesuai dengan kalimat yang ingin diucapkan. Ada beberapa alternatif teknik yang dapat digunakan untuk implementasi bagian ini. Dua teknik yang banyak digunakan adalah *formant synthesizer*, serta *diphone concatenation* [7].

Teknologi semakin pesat dan ditujukan untuk membantu kehidupan manusia. Perangkat lunak *text-to-speech* merupakan salah satu di antaranya, yang bertujuan untuk mengubah teks menjadi suara pengucapannya. Perangkat lunak ini dibuat dengan menggunakan metode *Syllable Concatenation*, yang pada dasarnya memainkan dan menggabungkan beberapa file suara dari suku kata-suku kata yang ditemukan di dalam input teks dan menggunakan bahasa pemrograman C++, dengan kompilator Microsoft Visual C++ 6.0. Selain perangkat lunak utama tersebut, juga dibuat beberapa perangkat lunak penunjang seperti: perangkat lunak perekaman suku kata, pengaturan singkatan, akronim dan diftong, dan pengaturan suku kata.

Perangkat lunak ini dapat dimanfaatkan untuk membantu mereka yang ingin belajar pengucapan kata dalam bahasa Indonesia dan Inggris serta berbagai tujuan lainnya. Perangkat lunak ini diujikan terhadap beberapa variasi input teks. Dan hasil pengujian menunjukkan bahwa proses normalisasi, pemenggalan kata dan penghasilan suara yang dilakukan sudah cukup baik. Kualitas suara yang dihasilkan oleh perangkat lunak utama ini bergantung pada kualitas suara masing-masing suku kata yang direkam sebelumnya [8].

Salah satu penelitian yang mendukung penelitian ini adalah Paper yang ditulis oleh Nakagawa. Penelitian ini didesain pembacaan teks yang ditulis oleh tangan dalam bahasa Jepang. Tulisan Jepang terdiri dari huruf katakana, hiragana, dan kanji karena itu dalam pembuatannya didesain database yang berjumlah 3 milion pola dari 283 partisipan. Metode ini menggunakan Metode Statistik dengan cara dicari tingkat kesamaan tertinggi antara teks yang dicari dengan database yang telah dibuat. Penelitian ini masih sebatas mengubah informasi image (yang berupa tulisan tangan) menjadi informasi teks [9].



Gambar 1: Gambaran konversi informasi

Penelitian lainnya yang berkaitan dengan penelitian *text to sound* adalah penelitian yang dilakukan oleh Srihari di Center of Excellence for Document Analysis and Recognition (CEDAR) University at Buffalo, State University of New York, Buffalo, USA. Paper ini membahas tentang pembacaan tulisan tangan yang discan. Metode yang digunakan adalah indexed menggunakan global image features, stroke width, slant, dan word gaps, dimana dengan metode-metode ini (hybrid) karakter yang ada pada citra akan dikenali. Sistem ini dibangun menggunakan Microsoft Visual C++ dan a relational database system. Penelitian ini masih sebatas menginterpretasi tulisan tangan (citra) dengan keluaran informasi teks [10].

Word Image Retrieval Using Binary Feature yang ditulis oleh Bin Zhang adalah penelitian yang membaca tulisan tangan dengan menggunakan Metode Binay Feature. Masing-masing kata diindeks yang selanjutnya digunakan Metode Gradient-based Binary Feature dan dihybrid dengan Dynamic Time Warping (DTW) untuk mencocokkan karakter yang diekstrak dengan database yang telah dibuat [11].

III. METODE PENELITIAN

Preprosesing berfungsi untuk mengkondisikan citra (*image*) sehingga dapat ditemukan posisi karakter dan memisahkan karakter dengan latar belakangnya. Sehingga citra teks dapat diproses pada tahap selanjutnya. Feature extraction adalah mengambil informasi citra dalam kasus ini adalah teks yang berada pada citra keseluruhan dan mengubahnya menjadi informasi vektor. Normalisasi teks adalah proses untuk mengubah ukuran teks dijadikan ukuran standart. Jadi meskipun ukuran teks yang ada pada citra beragam untuk mempermudah proses semua ukuran dinormalisasi menjadi satu ukuran. Segmentasi karakter adalah proses memisahkan karakter per karakter. Tujuan dari proses ini adalah untuk mempermudah pengenalan tiap karakter pada citra. Normalisasi karakter semua karakter yang mempunyai ukuran yang berbeda-beda, distandarisasi ukuranya sehingga didapatkan karakter yang mempunyai ukuran yang sama. Karakter yang sudah dinormalisasi dikenali karakter per karakter dengan menggunakan metode SOM Kohonen (Unsupervised Artificial Intelence). Karakter yang telah dikenali diadakan pembelajaran agar dapat membaca teks dengan baik. Hasil pembacaan ini akan direpresentasikan dalam bentuk suara. Konversi informasi digambarkan pada Gambar 1.

Learning Karakter Topologi SOM-Kohonen

Pada jaringan *self organizing*, suatu lapisan yang berisi neuron-neuron akan menyusun dirinya sendiri berdasarkan input nilai tertentu dalam suatu kelompok yang dikenal dengan istilah *cluster*. Selama proses penyusunan diri, *cluster* yang memiliki vektor bobot paling cocok dengan pola input (memiliki jarak yang paling dekat) akan terpilih sebagai pemenang. *Neuron* yang menjadi pemenang beserta neuron-neuron tetangganya akan memperbaiki bobot-bobotnya. Apabila kita ingin membagi data-data menjadi *K cluster*, maka lapisan kompetitif akan terdiri atas *K* buah neuron.

Learning karakter menggunakan SOM Kohonen dengan tahapan sebagai berikut:

1. Inisialisasi input
2. Penentuan jumlah kelas, dalam penelitian ini digunakan 36 kelas yang terdiri dari karakter A sampai Z dan karakter 0 sampai 9, tiap kelas terdiri dari 3 karakter yang berbeda.
3. Pengesetan parameter learning rate atau α , learning rate yang digunakan adalah $0,9 (1 - t/1000)$ dan $0,4/t$.
4. Nilai iterasi dilakukan sebanyak 1000 kali
5. Pemilihan nilai acak dari range input image berukuran: 10×20 pixel
6. Menghitung jarak antara data acak ke setiap bobot input, ke - *i* (D_i) dengan rumus :

$$D_i = \sqrt{\sum_{j=1}^m (W_{ij} - P_{zj})^2} \tag{1}$$

7. Pada langkah berikutnya adalah dengan Update bobot yang ke neuron ke *idx* :

$$W_{Idx-j} = W_{idx-j} + \alpha (P_{zj} - W_{idx-j}) \tag{2}$$

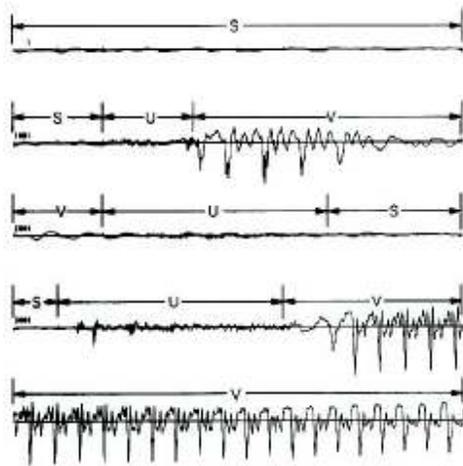
Tahapan-tahapan kerja di atas akan diulang sebanyak jumlah iterasinya yang telah ditentukan terlebih dulu.

Pengenalan Karakter

Pengenalan karakter menggunakan SOM-Kohonen berdasarkan bobot learning karakter yang telah dilakukan sebelumnya. Tahapannya adalah sebagai berikut:

1. Inisialisasi input
2. Pengesetan parameter learning rate atau α , learning rate yang digunakan adalah $0,9 (1 - t/1000)$
3. Penentuan kelas dengan mencari jarak terpendek, ke - *i* (D_i) dengan rumus :

$$D_i = \sqrt{\sum_{j=1}^m (W_{ij} - P_{zj})^2} \tag{3}$$



Gambar 2: Contoh Sinyal Ucapan "It's time"

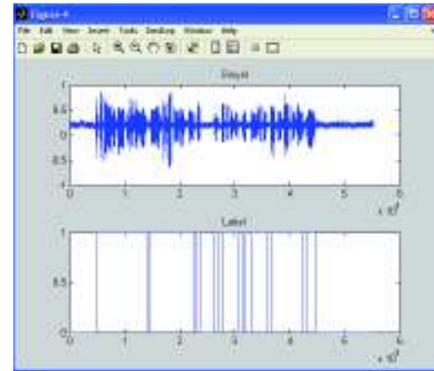
Skenario Machine Learning

Skenario dari pembacaan ini adalah dari karakter yang telah dikenali dibaca hurufnya yang selanjutnya disajikan dalam bentuk suara. Sinyal ucapan merupakan sinyal yang berubah terhadap waktu dengan kecepatan perubahan yang relatif lambat. Jika diamati pada selang waktu yang pendek (antara 5 sampai dengan 100 mili detik), karakteristiknya praktis bersifat tetap; tetapi jika diamati pada selang waktu yang lebih panjang karakteristiknya terlihat berubah-ubah sesuai dengan kalimat yang sedang diucapkan. Gambar 2 memperlihatkan contoh sinyal ucapan dari suatu kalimat bahasa Inggris "It's time" yang diucapkan oleh seorang pria. Setiap baris pada gambar tersebut memperlihatkan potongan sinyal selama 100 mili detik, sehingga seluruh gambar tersebut memperlihatkan sinyal ucapan sepanjang 500 mili detik.

Ada berbagai cara untuk mengklasifikasikan bagian-bagian atau komponen sinyal ucapan. Salah satu cara yang sederhana adalah dengan cara mengklasifikasikannya menjadi tiga keadaan yang berbeda, yaitu (1) *silence* (S), keadaan pada saat tidak ada ucapan yang diucapkan; (2) *unvoiced* (U), keadaan pada saat *vocal cord* tidak melakukan vibrasi, sehingga suara yang dihasilkan bersifat tidak periodik atau bersifat random; (3) *voiced* (V), keadaan pada saat terjadinya vibrasi pada *vocal cord*, sehingga menghasilkan suara yang bersifat kuasi periodik.

Pada Gambar 2 sudah tercantum label-label S, U dan V yang dapat mempermudah untuk mengamati perbedaan keadaan-keadaan tersebut. Baris pertama serta awal baris kedua ditandai dengan S, artinya bagian tersebut merepresentasikan keadaan diam dimana pembicara belum mengucapkan apapun. Amplitudo kecil yang tampak pada perioda tersebut merupakan noise latar belakang yang ikut terekam.

Suatu perioda singkat *unvoiced* (U) tampak mendahului vocal pertama dalam kata "It". Selanjutnya diikuti oleh daerah *voiced* (V) yang cukup panjang, merepresentasikan vokal "i".



Gambar 3: Proses labelling untuk pencarian feature

Berikutnya diikuti oleh daerah *unvoiced* (U) yang merepresentasikan daerah pelemahan pengucapan "i". Setelah itu diikuti oleh *silence* (S) yang merupakan bagian dari fonem "t", dan seterusnya.

Dari contoh tersebut jelas bahwa segmentasi ucapan menjadi S, U dan V tidak bersifat eksak, artinya ada daerah-daerah yang tidak dapat dikategorikan dengan tegas ke dalam salah satu dari tiga kategori tersebut. Salah satu penyebabnya adalah perubahan dari keadaan-keadaan alat ucap manusia yang tidak bersifat diskrit dari satu keadaan ke keadaan lainnya, sehingga bunyi transisi dari satu segmen ke segmen lainnya menghasilkan bentuk yang tidak mudah ditentukan. Selain itu, ada segmen-segmen ucapan yang mirip atau bahkan mengandung *silence* didalamnya.

IV. ANALISIS DATA DAN PEMBAHASAN

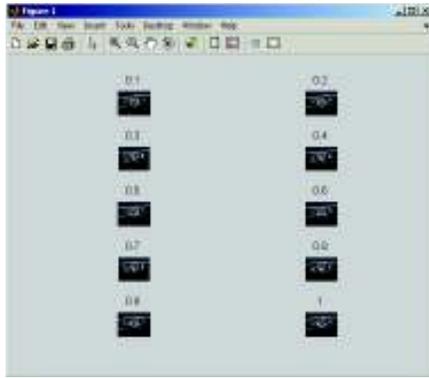
A. Preprocessing

Proses preprocessing terdiri dari: Image diresize menjadi 0,6 kali dari besar *image* semula, proses greyscale, edge detection, menyambung pixel yang mempunyai jarak kurang dari sama dengan 10 pixel, menghilangkan noise, mengindex area yang mempunyai kemungkinan merupakan objek atau plat nomer, mencari luas area yang telah diindex, mencari panjang area yang telah diindex, mencari tinggi area yang telah diindex dan mencari perbandingan antara tinggi dan panjang area yang telah diindex.

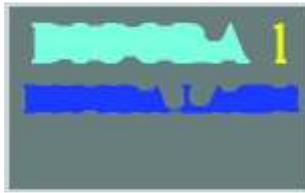
Tahapan proses: image diresize 0,6 dari ukuran semula, image berwarna dijadikan image *greyscale*, *Edge Detection Sobel*, menyambung pixel yang berjarak 10 pixel (*imclose*), menghilangkan hole/lubang yang ada, mencari garis tepi tiap objek pada image, mencari luas tiap area.

B. Segmentasi Kata dalam kalimat

Metode yang digunakan untuk mensegmentasi kata dalam kalimat adalah Metode Index. Proses Metode Index adalah hasil preprocessing diindex sehingga dihasilkan kandidat kata



Gambar 4: Image diresize menjadi 0,1 sampai 1 kali dari besar image semula



Gambar 5: Index kata dalam kalimat

dalam kalimat. Sehingga segmentasi per kata dalam kalimat dapat dilakukan. Contoh proses ini digambarkan pada Gambar 5. Hasil dari proses akhirnya dijadikan informasi untuk cropping kata dalam kalimat.

Logika Metode Index adalah dikembangkannya pixel per karakter sebesar 10 pixel sehingga dapat menggandeng karakter selanjutnya dalam sebuah kata. Bila kalimat terdapat spasi maka kata satu dengan kata selanjutnya tidak akan terhubung sehingga kata dalam sebuah kalimat akan terpisah.

C. Normalisasi kata dalam kalimat

Normalisasi kata yang digunakan metode perbandingan. Ukuran kata yang telah dicrop mempunyai ukuran yang beraneka ragam sehingga ukuran harus dinormalisasi agar dihasilkan ukuran plat yang standart 600 x 1000 pixel. Gambar 8 adalah contoh dari hasil normalisasi kata dalam kalimat. Tujuan dari normalisasi kata dalam kalimat adalah agar seluruh kata ukurannya menjadi normal atau standart. Proses ini menghindari ukuran kata yang terlalu besar atau terlalu kecil. Proses ini juga mempermudah untuk segmentasi kalimat pada proses selanjutnya.



Gambar 6: Hasil segmentasi kata pertama



Gambar 7: Hasil segmentasi kata pertama dan diproses selanjutnya



Gambar 8: Normalisasi kata dalam kalimat

Ukuran hasil cropping adalah (gx,gy) akan dijadikan ukuran kata dalam kalimat yang standart 600 x 1000 pixel, sehingga mempunyai skala:

$$Skalax = 600/gx$$

$$Skalay = 1000/gy$$

Sehingga pembesaran plat normalisasi adalah:

$$gx' = gx \times skalax$$

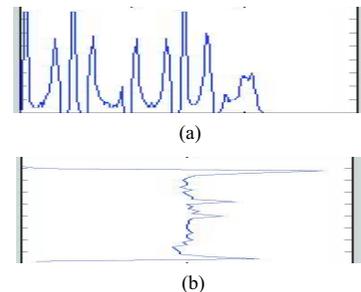
$$gy' = gy \times skalay$$

D. Segmentasi Karakter pada Kata

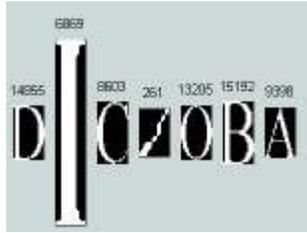
Metode untuk mensegmentasi karakter dalam kata adalah Coulomb Sum Graph dan Line Sum Graph, serta Indeks kandidat karakter.

Coulomb Sum Graph dan Line Sum Graph

Gambar 9.a adalah Line Sum Graph dari kalimat dan Gambar 9.b adalah Coulomb Sum Graph. Metode ini dapat diterapkan bila ukuran kata atau karakter mempunyai ukuran yang sama sedangkan pada kenyataannya ukuran karakter pada kalimat yang discan mempunyai ukuran yang tidak sama.



Gambar 9: (a).Line Sum Graph, (b).Crop Coulomb Sum Graph



Gambar 10: Segmentasi Karakter



Gambar 11: Crop gambar threshold

Indeks kandidat karakter

Tiap area diindeks dan dicari luas dan perbandingan lebar dan panjangnya. Dari proses statistik perbandingan lebar dan panjang karakter adalah 1 sampai 5 sehingga tiap area kandidat difilter. Hasil dari filter di filter kembali dengan menggunakan luasan yang memenuhi standart. Luasan yang memenuhi standar adalah 2500 pixel sampai 15000 pixel. Gambar 10 adalah contoh dari proses segmentasi karakter dengan menggunakan metodek index dan filter morphology.

E. Normalisasi Karakter

Seluruh karakter hasil dari segmentasi dinormalisasi menjadi ukuran 10 x 20 pixel. Hal ini dilakukan untuk mempermudah pengenalan karakter. Gambar 11 adalah contoh dari segmentasi karakter.

Normalisasi mempunyai skala:

$$Skalaxi = 10/gxi$$

$$Skalayi = 20/gyi$$

Sehingga pembesaran plat normalisasi adalah:



Gambar 12: Karakter yang dilearning

$$gxi' = gxi \times skalax$$

$$gyi' = gyi \times skalay$$

Ukuran standart dari plat yang telah dinormalisasi adalah (gxi',gyi') atau 10 x 20 pixel

F. Learning Karakter

Learning karakter dilakukan dengan menggunakan SOM Kohonen. Penelitian ini dipilih 36 kelas (cluster), dimana tiap kelas terdiri dari 2 karakter dengan ukuran 10 x 20 pixel. Karakter yang dilearning dalam penelitian digambarkan pada Gambar 12.

Proses klasifikasi menggunakan tahapan sebagai berikut:

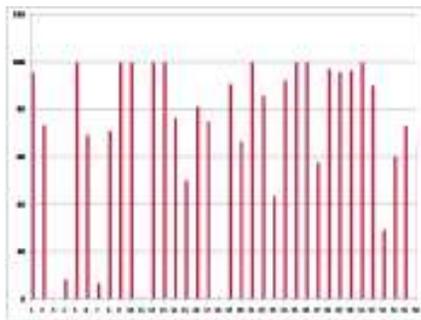
1. Inisialisasi input
 Penelitian ini data input tidak dilakukan inialisasi sebab data yang dimasukkan adalah data biner yang terdiri dari nilai 0 dan 1.
2. Penentuan jumlah kelas, dalam penelitian ini digunakan 36 kelas yang terdiri dari karakter A sampai Z dan karakter 0 sampai 9, tiap kelas terdiri dari 2 karakter yang berbeda.
3. Penentuan bobot awal tiap kelas atau cluster.
4. Pengesetan parameter learning rate atau α , learning rate yang digunakan adalah $0,9 (1 - t/1000)$ dan optimum learning rate $\alpha = A/(t+B)$, dima A dan B adalah konstanta.
5. Nilai iterasi dilakukan sebanyak 1000 kali untuk learning rate $0,9 (1 - t/1000)$ dan 100 untuk optimum learning rate.
6. Pemilihan nilai acak dari range input image berukuran : 10 x 20 pixel
7. Menghitung jarak antara data acak ke setiap bobot input, ke - i (D_i) dengan rumus : $D_i = \text{argmin} \|W - P\|$
 Bila diturunkan menjadi

$$D_i = \sqrt{\sum_{j=1}^m (W_{ij} - P_{zj})^2} \tag{4}$$

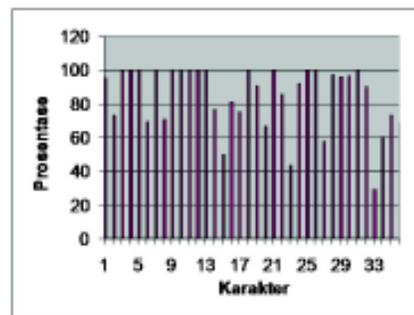
8. Pada langkah berikutnya adalah dengan Update bobot yang ke neuron ke idx :

$$W_{Idx-j} = W_{idx-j} + \alpha (P_{zj} - W_{idx-j}) \tag{5}$$

Bobot yang dihasilkan oleh learning karakter digunakan untuk mengenali karakter pada plat mobil.



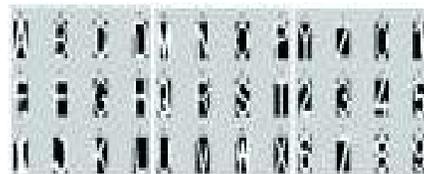
Gambar 13: Prosentase keberhasilan pengenalan per karakter dengan menggunakan learning rate 0,9 ($1 - t/1000$) dan jumlah iterasi 1000.



Gambar 15: Prosentase keberhasilan pengenalan per karakter dengan menggunakan learning rate 0,4/t dan jumlah iterasi 100.



(a) learning rate 0,9($1 - t/1000$), jumlah iterasi 1000



(b) learning rate 0,4/t, dengan jumlah iterasi 100

Gambar 14: Visualisasi bobot per karakter

V. HASIL

Berdasarkan Gambar 13 prosentase tingkat keberhasilan pembacaan karakter bervariasi, hal ini dipengaruhi oleh bobot yang dihasilkan ketika proses learning. Karakter yang dilearning adalah Gambar 12 dihasilkan bobot akhir yang dapat divisualisasikan pada Gambar 14.

Gambar 14 memvisualisasikan evolusi karakter yang terjadi, hal ini mempengaruhi tingkat keberhasilan pembacaan suatu karakter. Diantara karakter yang mengalami evolusi seperti ditunjukkan oleh Tabel I.

Berdasarkan Tabel II maka dapat disimpulkan bahwa tingkat akurasi pembacaan karakter dengan menggunakan learning rate $0,9(1 - t/1000)$ dengan jumlah iterasi 1000 dapat diperbaiki bila menggunakan optimum learning rate 0,4/t dengan jumlah iterasi 100. Tingkat akurasi pembacaan karakter menggunakan learning rate 0,4/t digambarkan pada Gambar 15. Berdasarkan Gambar 15 penggunaan optimum learning rate 0,4/t dapat memperbaiki tingkat akurasi pembacaan karakter tetapi belum memperbaiki keseluruhan karakter. Misalnya karakter 7 terbaca 7 terkadang terbaca sebagai karakter Z.

Dalam teknologi musik, sound syntesis adalah proses untuk memproduksi suara dari analog ke digital. Proses men-

generate, menggabungkan atau mencampur suara adalah suatu proses yang kompleks. Sound syntesis digunakan di dunia musik, computer, ilmu pengetahuan, film, akustik dan biologi.. Text to Speech (TTS) adalah suatu sistem pengkonversi teks menjadi suara atau ucapan manusia. Akan tetapi pada teknologi Text to Speech, ada beberapa kendala teknis yang harus diatasi untuk bisa meningkatkan kualitas suara yang dihasilkan, salah satunya adalah dengan penambahan kosakata yang dapat dikenali oleh sistem serta kata atau kalimat yang diucapkan bisa memiliki prosodi atau intonasi. Untuk itu diperlukan cara bagaimana penambahan kosakata dan pengucapan kalimat yang memiliki prosodi pada sistem TTS dapat dilakukan.

Sistem Text to Speech secara garis besar terdiri dari dua sub sistem utama, yaitu: NLP (Natural Language Processing) atau Text to Phoneme dan DSP (Digital Signal Processing) atau Phoneme to Speech. Text to Phoneme dari sistem TTS ini yang telah dibuat, dimana pada bagian ini dibuat GUI (Grafik User Interface) dari sistem TTS dan algoritma pemrograman untuk: mengenali teks, membuat kamus untuk kata-kata khusus serta bagaimana cara agar prosodi suara dapat dihasilkan. Sedangkan pada bagian Phoneme to Speech, sub sistem TTS ditangani oleh aplikasi MBROLA yang digunakan sebagai speech synthesizer. Text to Speech bahasa Indonesia dengan model prosodi dataset pada MBROLA, yang dapat mengetahui cara meningkatkan kualitas suara yang dihasilkan untuk pengembangan aplikasi selanjutnya.

Hasil yang didapatkan yaitu system yang dibangun paling baik bekerja untuk system yang mempunyai besar huruf antara 10 poin sampai 20 poin. Bila yang diujikan mempunyai ukuran diluar diatas masih dapat dibaca tetapi tidak dengan akurat.

VI. SIMPULAN

Sistem yang dibangun dengan menggunakan metode SOM Kohonen beserta serangkaian proses preprocessing telah dapat mengubah informasi image karakter menjadi teks dengan tingkat akurasi tertentu tiap karakter. Dalam proses teks to sound masih mempunyai keterbatasan hanya ukuran karakter tertentu saja yang dapat dibaca dengan tepat.

TABEL I: Visualisasi bobot yang terbaca sebagai karakter lain

No	Karakter	Terbaca sebagai	Analisa
1.	C	G	Ketika learning karakter G ada yang masuk pada kelas C
2.	D	C	Ketika learning karakter C ada yang masuk pada kelas D
3.	F	F dan P	Ketika learning karakter P ada yang masuk pada kelas F
3.	G	O	Ketika learning karakter O ada yang masuk pada kelas G
4.	K	W	Ketika learning karakter W ada yang masuk pada kelas K
5.	O	D	Ketika learning karakter D ada yang masuk pada kelas O
6.	P	R	Ketika learning karakter R ada yang masuk pada kelas P

TABEL II: Tabel keterbacaan karakter

Karakter	Gambar 14	Terbaca dengan learning rate	
		$0,4/t$	$0,9(1 - t/1000)$
K	(a)	K	R
C	(b)	C	D
P	(c)	P	F
D	(d) (h)	D	O
G	(e)	G	C
R	(f)	R	P
W	(g)	W	K

[1] <http://www.freelists.org/post/ppi/ppiindia-15-Juta-Penduduk-Indonesia-Buta-Huruf>

[2] <http://www.menkokesra.go.id/content/view/5236/39/>

[3] <http://www.detikinet.com/index.php/detik.read/tahun/2005/bulan/08/tgl/03/time/124151/idnews/415139/idkanal/398>

[4] <http://www.alatperagaluarbiasa.com/>

[5] Octoni Simbolon, *TEXT-TO-SPEECH Indonesia dengan Menggunakan MBROLA SYNTHESIZER*, Gunadarma, 2007.

[6] Arry Akhmad, *EXT-TO-SPEECH Indonesia dengan Menggunakan MBROLA SYNTHESIZER*, ITB, 2002

[7] <http://www.amirfaisal.web.ugm.ac.id/2008/09/14/indotts-indonesian-text-to-speech/>

[8] <http://dewey.petra.ac.id/dgt-res-detail.php?mode=extended&knokat=7589>

[9] Nakagawa, *Collection of on-line Handwritten Japanese Character Pattern Databases and Their Analyses*, International Journal on Document Analysis and Recognition, 2004

[10] Sargur Srihari, Chen Huang and Harish Srinivasan, *Content-based Information Retrieval from handwritten Documents*, Center of Excellence for Document Analysis and Recognition (CEDAR), University at Buffalo, State University of New York, Buffalo, USA, 2002.

[11] Zhang, *Word Image Retrieval Using Binary Feature*, Center of Excellence for Document Analysis and Recognition (CEDAR), University at Buffalo, State University of New York, Buffalo, USA, 2004.