

Identifikasi Biometrik Intonasi Suara untuk Sistem Keamanan Berbasis Mikrokomputer

I Nyoman Kusuma Wardana¹⁾, I Gede Harsemadi²⁾

^{1,2)}STMIK STIKOM Bali

Jl. Raya Puputan Renon No.86 Denpasar, telp. +62(361)244445

e-mail: kusumawardana@stikom-bali.ac.id¹⁾, harsemadi@stikom-bali.ac.id²⁾

Abstrak

Pada penelitian ini, akan dicoba melakukan pengenalan intonasi sebagai langkah pengamanan ganda pada kata sandi berbasis suara. Pola kontur intonasi akan dikenali menggunakan jaringan saraf tiruan (JST), sedangkan pengenalan suara dilakukan dengan pengukuran jarak Euclidean (Euclidean distance) dengan sistem pemrograman dinamis (dynamic programming). Dalam penerapannya, selain jenis kata yang dimasukkan, sistem juga akan mengecek alur intonasi dari input suara tersebut. Pada penelitian ini, akan dilakukan pemrosesan sinyal suara berbasis mikrokomputer dan mikrokontroler. Kombinasi kedua jenis alat ini dapat dijadikan solusi selain penggunaan komputer desktop atau laptop, terutama dari efisiensi komputasi dan ukuran alat yang jauh lebih kecil. Mikrokontroler dan mikrokomputer memiliki fitur yang cukup terbatas jika dibandingkan dengan komputer desktop, baik dari segi memori program, memori data, port input/output dan sebagainya. Luaran pada penelitian ini adalah terciptanya alat keamanan berbasis kata sandi suara dengan sistem pengamanan ganda untuk akses keamanan pintu Laboratorium Hardware STMIK STIKOM Bali. Hasil penelitian menunjukkan bahwa pengenalan cenderung kearah speaker dependent dengan tingkat keberhasilan 81,3%.

Kata kunci: pitch, intonasi, mikrokomputer, mikrokontroler, dynamic programming

Abstract

In this research, the authors tried to identify the intonations following the uttered words as a double protection system. The pitch contours are recognized using artificial neural networks, whereas the uttered words are matched by means of Euclidean distance (based on Dynamic Programming techniques). In its applications, both kind of words and its intonations are simultaneously identified. In this research, the speech processing systems are based on microcomputer and microcontroller. By using these two devices, the use of desktop or laptop computers can be avoided. Microcomputers and microcontrollers are small-sized devices and can be applied for embedded systems purposes. They have limited features, such as memories, I/O ports, etc. Therefore, the efficiency of programming structure is strictly needed. The research results was a device that can be used as a double protections systems and it can be attached in laboratory's door of STMIK STIKOM Bali. Based on the research, the system was tended to speaker dependent system with 81.3% of achievement.

Keywords: pitch, intonation, microcomputer, microcontroller, dynamic programming

1. Pendahuluan

Komunikasi melalui suara merupakan salah satu kemampuan dasar terpenting yang dimiliki oleh manusia. Manusia berkomunikasi dengan sesama melalui proses berbicara dan mendengar untuk menangkap berbagai informasi dalam proses komunikasi tersebut. Usaha peningkatan kemampuan komunikasi semakin giat dikembangkan, dan impian untuk berkomunikasi dengan siapapun, dimanapun, dan kapanpun semakin menjadi kenyataan. Impian ini terus berlanjut dan meluas, sehingga menuntut teknologi interaksi mesin dan manusia (*human-machine interfaces*) yang semakin canggih dan handal. Dampak nyata dari perkembangan ini terlihat pada semakin canggihnya teknologi pemrosesan sinyal digital, sintesis suara, ataupun sistem pengenalan suara (*digital speech processing, synthesis and recognition technology*). Selama proses komunikasi berlangsung, gelombang suara membawa informasi linguistik, karakter vokal pembicara ataupun emosi pembicara. Informasi tersebut dapat diperoleh melalui tinjauan dan analisa secara fisik terhadap gelombang suara baik dalam domain waktu ataupun domain frekuensi.

Salah satu parameter linguistik yang penting untuk dianalisis adalah **intonasi** suara. Melalui perbedaan intonasi, seorang pendengar dapat menerima informasi dan berinteraksi dengan pembicara. Intonasi perintah, pertanyaan, dan intonasi pernyataan merupakan intonasi-intonasi dasar yang

membangun sistem interaksi baik antar manusia maupun antar mesin dan manusia (*human-machine interfaces*). Contoh nyata dari sistem interaksi mesin dan manusia adalah perkembangan robot cerdas yang terkendali suara. Dengan intonasi tertentu, robot dapat memahami apakah diperintah untuk melakukan sesuatu ataupun ditanya untuk menjawab suatu hal.

Tantangan besar yang dihadapi dalam sistem komunikasi antara mesin dan manusia adalah kemampuan komputer untuk memahami suatu karakter suara dari pembicara dengan berbagai latar belakang budaya, bahasa ibu ataupun jenis kelamin pembicara. Hal ini penting, terutama untuk sistem keamanan seperti *password* suara dan sebagainya. Demikian pula pengenalan intonasi dapat bersifat khusus untuk berbagai latar belakang budaya, bahasa ibu ataupun jenis kelamin pembicara.

Teknologi pengolahan suara dapat melahirkan berbagai penerapan, diantaranya kata sandi (*password*) suara untuk sistem keamanan. Umumnya, akses menggunakan kata sandi berbasis suara akan diterima jika memenuhi kondisi sebagai berikut:

- 1) kata yang diinputkan sesuai dengan data yang tersimpan pada sistem, dan
- 2) pembicara (*talker*) yang melakukan input suara telah terdata secara resmi pada sistem (bersifat *speaker dependent*).

Kondisi pertama memiliki peluang untuk diketahui oleh pihak lain yang tidak memiliki akses, sedangkan kondisi kedua dapat dijadikan sebagai benteng kedua. Namun demikian, kondisi kedua pun dapat terganggu, terutama jika seseorang sedang memiliki kendala pada sistem produksi suara.

Pada penelitian ini, akan dicoba melakukan pengenalan intonasi sebagai langkah pengamanan ganda pada kata sandi berbasis suara. Pola kontur intonasi akan dikenali menggunakan jaringan saraf tiruan (JST), sedangkan pengenalan suara dilakukan dengan pengukuran jarak Euclidean (*Euclidean distance*) dengan sistem pemrograman dinamis (*Dynamic Programming*). Dalam penerapannya, selain jenis kata yang dimasukkan, sistem juga akan mengecek alur intonasi dari input suara tersebut. Parameter intonasi suara cukup unik, dan bersifat tahan terhadap gangguan sistem produksi suara, misalnya dalam kondisi serak. Pada penelitian ini, akan dilakukan pemrosesan sinyal suara berbasis mikrokomputer dan mikrokontroler. Kombinasi kedua jenis alat ini dapat dijadikan solusi selain penggunaan komputer desktop atau laptop, terutama dari efisiensi komputasi dan ukuran alat yang jauh lebih kecil.

Luaran yang diharapkan setelah selesai dilakukannya penelitian ini adalah terciptanya alat keamanan berbasis kata sandi suara dengan sistem pengamanan ganda, yaitu berupa jenis kata dan intonasi yang mengiringinya. Dibandingkan dengan penggunaan komputer desktop atau laptop untuk melakukan proses pengolahan suara, alat yang akan tercipta akan berukuran jauh lebih kecil karena menggunakan mikrokomputer, bersifat hemat energi dan dapat dikemas dengan baik sehingga dapat dikomersialkan.

2. Tinjauan Pustaka

Lupu dan Cioban (2009) membagi sistem keamanan terhadap suatu akses tertentu menjadi tiga kategori utama, yaitu objek (misalnya penggunaan kunci), informasi (misalnya penggunaan kode), dan karakteristik personal/biometrik (misalnya sidik jari dan suara). Penerapan suara dapat dibagi lagi menjadi tiga kategori utama, yaitu penerapan di bidang telekomunikasi, mengetahui kehadiran secara fisik, dan untuk keperluan forensik [1].

Bidang keilmuan yang banyak diterapkan pada sistem pengenalan suara adalah *automatic speech recognition* (ASR). Ruang lingkup ASR mencakup berbagai tahapan dalam mendesain suatu sistem pengenalan suara. Kriteria ini mencakup *pre-processing filter*, *end-point detection*, *feature extraction techniques*, *speech classifier*, *database*, dan *performance evaluation* [2]. Dari berbagai proses tersebut, umumnya terdapat beberapa metode yang umum digunakan, yaitu spektogram, *dynamic time warping*, *vector quantization*, jaringan saraf tiruan (JST), hidden Markov model (HMM), dan metode statistik [1].

Khusus untuk sistem keamanan, terdapat berbagai metode keamanan yang banyak digunakan, seperti memasukkan pin tertentu, penggunaan kunci (baik digital maupun elektronik), pemanfaatan kartu identitas, kriptografi, pengenalan ciri-ciri biometrik dan suara. Peran suara memiliki posisi tersendiri. Suara seseorang tidak dapat dicuri, hilang, terlupakan, ataupun ditirukan dengan tingkat akurasi yang tinggi [3]. Syazilawati dkk. (2006) dan Wahyudi (2007) telah melakukan penelitian dengan memanfaatkan suara sebagai sistem akses pintu, masing-masing memanfaatkan kemampuan jaringan saraf tiruan (JST) serta *Adaptive-Network Fuzzy Inference System* (ANFIS) [3],[4].

Berbagai pengetahuan untuk menyelesaikan masalah pengenalan suara telah dibentuk dengan pendekatan sistem kecerdasan buatan (*artificial intelligence*). Terdapat dua konsep penting dalam sistem kecerdasan buatan, yaitu sebagai berikut [5] :

1. Pembelajaran (*learning*).
2. Adaptasi (*adaptation*).

Kedua konsep ini menjadi dasar dalam mengimplementasikan jaringan saraf tiruan (JST) atau dikenal dengan istilah *artificial neural network* (ANN). JST diterapkan pada berbagai aplikasi seperti peramalan, dan pengenalan pola (*pattern recognition*), termasuk didalamnya pengenalan suara.

Jaringan saraf tiruan dan pengolahan suara (*speech processing*) merupakan dua disiplin ilmu yang banyak dijadikan sebagai area penelitian. Dua disiplin ilmu ini mempunyai sejarah perkembangan selama puluhan tahun, dan melibatkan berbagai disiplin ilmu lain seperti fisiologi, fisika, statistika, psikologi, linguistik dan teknik [6]. Pengolahan suara menjadi bidang yang sangat aktif diteliti bahkan sebelum era 1980-an dimana masa itu jaringan saraf tiruan kembali aktif diteliti [5].

Berbagai tipe sistem pengenalan telah diteliti, termasuk pendekatan secara statistik menggunakan metode *Hidden Markov Model* (HMM), penerapan *Support Vector Machines* (SVM) ataupun pemrograman dinamik menggunakan metode *Dinamic Time Warping* (DTW). Sistem hibrid juga telah dikembangkan, seperti hibrid HMM/SVM ataupun HMM/ANN [7]. Disisi lain, coding suara menggunakan teknik *vector quantitation* dan *autoregressive modeling* memegang peranan penting pengembangan sintesis suara dan sistem komunikasi. Pada saat yang sama, jaringan saraf tiruan memiliki keunggulan dalam hal kemampuan klasifikasi yang tinggi dan kedepan menjanjikan teknologi yang mutakhir untuk proses pengolahan suara [6].

Suatu ucapan dapat dibedakan apakah tergolong ucapan pernyataan (*declarative*) atau pertanyaan (*interrogative*) hanya melalui intonasinya [8]. Perbedaan intonasi akan menggambarkan perbedaan informasi yang disampaikan oleh pembicara. Ketika kita mendengar sebuah ucapan “kamu sudah makan”, maka kemungkinan pendengar akan menangkap sebuah ucapan pernyataan ataupun sebuah pertanyaan. Sudah pasti dengan intonasi berbeda akan menyebabkan tanggapan yang diterima oleh pendengar juga akan berbeda.

Pendeteksian intonasi telah dilakukan oleh beberapa peneliti, terutama untuk intonasi Bahasa Inggris. Suatu ucapan terdiri dari satu atau lebih bagian-bagian intonasi [8]. Parameter yang digunakan untuk melihat fenomena intonasi adalah *pitch*. Penelitian tentang *pitch* dilakukan dengan menganalisis kontur frekuensi fundamental (*F0 contours*) dari sebuah ucapan. Di sisi lain, pencarian *pitch* yang lebih teliti dilakukan dengan mempertimbangkan efek laringalisasi (*laryngealization effects*). Efek ini tergambar sebagai eksitasi suara yang tidak teratur. Hal ini disebabkan oleh sinyal suara yang memiliki periodisitas yang tidak reguler, memiliki variasi amplitudo yang besar, atau memiliki periode *pitch* yang terlalu lama [8].

Estimasi terhadap frekuensi fundamental atau yang dikenal dengan deteksi *pitch*, merupakan topik yang sangat populer untuk diteliti dalam beberapa tahun, bahkan sampai saat ini. Masalah dasar dalam mengekstraksi frekuensi fundamental (*fo*) dari suatu gelombang suara adalah pencarian bagian-bagian komponen gelombang yang memiliki frekuensi terendah. Suatu bagian-bagian (*partials*) gelombang akan terkait dengan bagian lainnya [9]. Bagian terendah (*first partial*) adalah frekuensi fundamental (*fo*) dari suatu gelombang suara.

Wardana (2008) telah membuat aplikasi untuk membedakan antara intonasi pernyataan (*declarative*) dan intonasi pertanyaan (*interrogative*) hanya melalui intonasinya untuk melakukan sistem kontrol dan monitoring penggunaan perangkat listrik [10]. Sistem yang dibangun menggunakan jaringan saraf tiruan untuk membedakan kedua intonasi tersebut, dan proses pengolahan suara memanfaatkan MATLAB pada aplikasi desktop. Namun demikian, penggunaan aplikasi desktop untuk beberapa kasus menjadi kurang efisien, sebagai contoh penerapan sistem keamanan untuk akses pintu. Penggunaan komputer desktop dirasa akan menghabiskan sumber daya listrik yang cukup besar, terutama jika sistem digunakan sepanjang hari. Oleh karena itu, pada penelitian ini, penulis mencoba mengganti aplikasi desktop dengan sebuah mikrokomputer, dan selanjutnya meneliti keandalan sistem.

3. Metode Penelitian

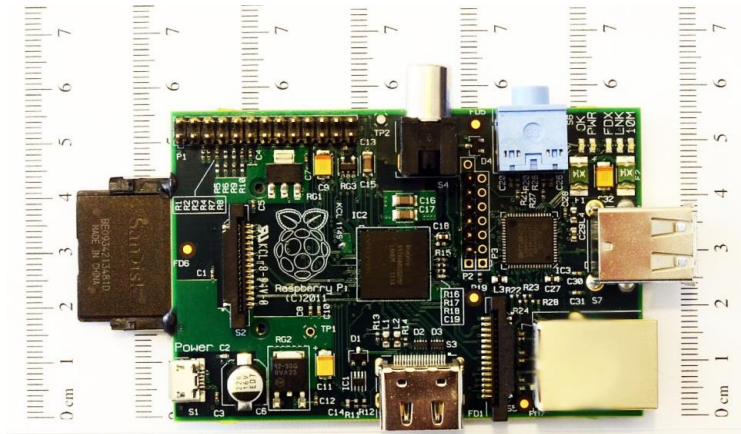
3.1. Tempat Penelitian

Penelitian akan dilakukan di Laboratorium Hardware, kampus STMIK STIKOM Bali, Jl. Raya Puputan No. 86 Renon, Denpasar-Bali

3.2. Rancangan Penelitian

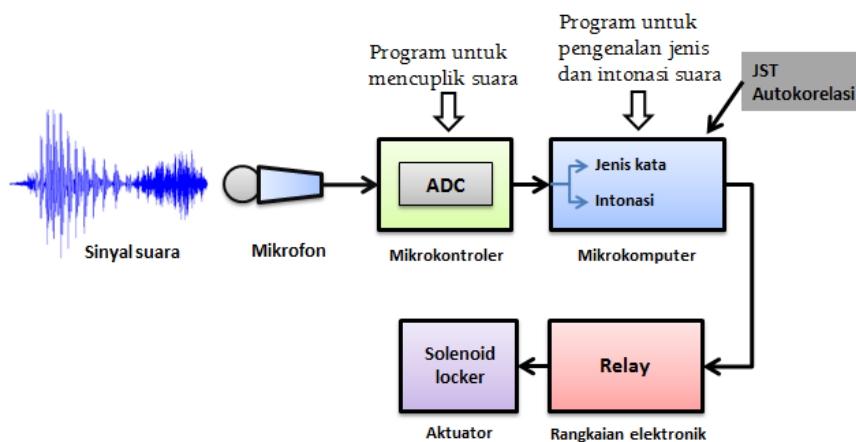
Mikrokomputer yang digunakan pada penelitian ini adalah Raspberry PI tipe B (Raspi B). Raspi B mempunyai arsitektur berbasis ARMv6, dengan RAM sebesar 512MB dan Prosesor dengan *clock speed* 700MHz. Untuk sistem operasi, sistem ini menggunakan *distro* Linux berbasis Debian yang di kstem *build* untuk arsitektur ARM dengan nama Raspbian, dan versi *build*-nya sendiri adalah Raspbian Wheezy. Daya yang dibutuhkan tidak begitu besar, yaitu sebesar 5 Volt tegangan searah (DC) dan konsumsi arus sebesar 700 mA. Raspi B sendiri telah dilengkapi berbagai antarmuka yang terdiri dari 8

General Purpose Input Output (GPIO) pin, USB 2.0 Port x 2, RJ-45 x 1, Audio jack 3.5 x 1, HDMI x 1, 1 slot Micro SD, dan konektor kamera dan LCD masing-masing berjumlah 1 buah [11]. Bentuk dan ukuran Raspi B diperlihatkan pada Gambar 1.



Gambar 1. Raspberry PI Tipe B

Dari sisi perangkat keras (*hardware*), penelitian ini mencoba menggabungkan mikrokontroler dan mikrokomputer, seperti yang ditunjukkan pada Gambar 2. Mikrokontroler akan bertugas untuk mencuplik input suara dari *user* dengan menggunakan mikrofon mini dan pengaturan waktu sampling yang tepat.

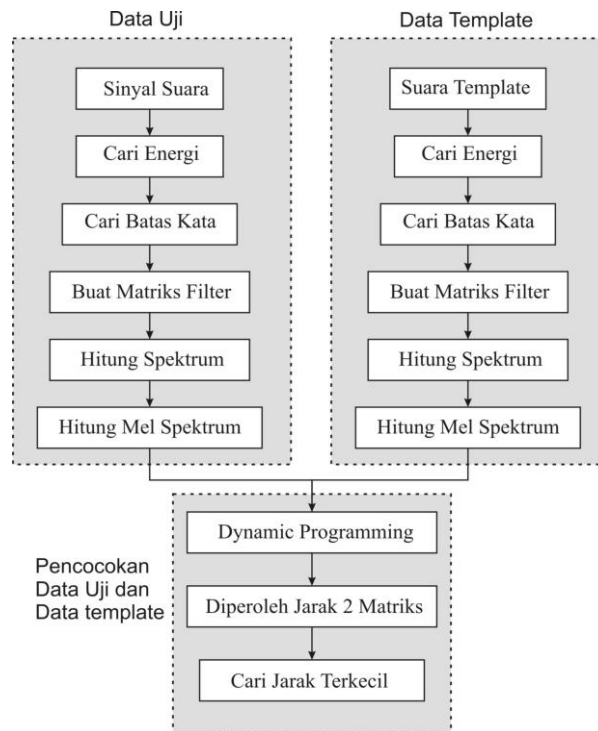


Gambar 2. Skema Keseluruhan Sistem

Mikrokontroler memiliki kemampuan yang andal untuk mencuplik sinyal analog (dalam hal ini gelombang suara) dengan memanfaatkan *Analog-to-Digital Converter* (ADC). Karena melibatkan komputasi yang lebih kompleks, proses pengolahan suara tidak tepat dilakukan pada mikrokontroler, namun selanjutnya akan diambil alih oleh mikrokomputer. Pada perancangan perangkat lunak (*software*), meliputi pemrograman mikrokontroler dan mikrokomputer.

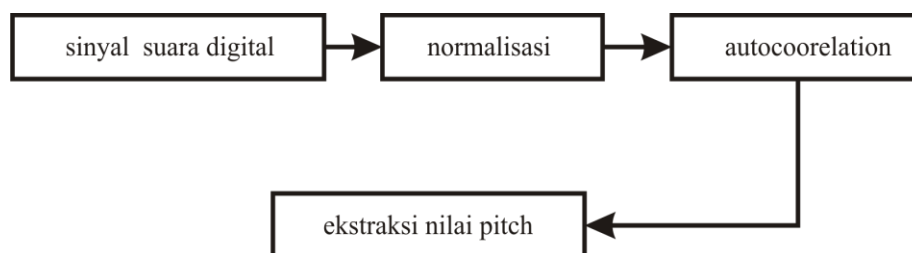
3.3. Metode Pengenalan Jenis Kata dan Intonasi

Metode yang digunakan untuk sistem pengenalan suara adalah pencarian jarak terdekat dua buah vektor. Dalam hal ini, akan dicari jarak terdekat dari suatu matriks mel suara input terhadap matriks mel suara *template*. Metode pencarian jarak terdekat memanfaatkan algoritma pengukuran jarak euclidean (*euclidean distance*) dengan sistem pemrograman dinamis (*Dynamic Programming*). Diagram alir untuk prngnalan jenis kata diperlihatkan seperti pada Gambar 3.



Gambar 3. Diagram Alir Sistem Pengenalan Suara

Program pengenalan intonasi dilakukan dengan mengekstrak nilai frekuensi fundamental (F0) dari suatu ucapan. Nilai F0 dianggap mewakili kontur intonasi (*pitch*). Metode yang digunakan untuk mencari frekuensi fundamental adalah metode autokorelasi dan jaringan saraf tiruan (JST). Metode autokorelasi diperlihatkan pada Gambar 4. Nilai-nilai frekuensi fundamental dari proses autokorelasi akan menjadi input untuk jaringan saraf tiruan, baik sebagai data latih maupun data uji. Agar sistem identifikasi terhadap jenis intonasi dapat dilakukan, maka sebelumnya dilakukan pembuatan *database* intonasi, ekstraksi nilai *pitch*, dan pemrograman jaringan saraf tiruan.



Gambar 4. Metode Autokorelasi

Jaringan saraf tiruan memerlukan data latih dan data uji berupa sampel suara. Berbagai sampel suara yang dilatih akan menentukan tingkat kecerdasan jaringan untuk mengenali pola-pola yang belum diketahuinya, dalam hal ini untuk membedakan suatu intonasi. Jaringan yang dibangun akan dicoba menggunakan input berupa nilai *pitch* suara. Melalui pemrograman, akan dicari metode pelatihan (*training*) yang paling tepat dalam pengenalan jaringan terhadap pola-pola input. Pencarian jumlah neuron lapisan tersembunyi juga akan dilakukan, agar *error* yang dihasilkan output seminimal mungkin.

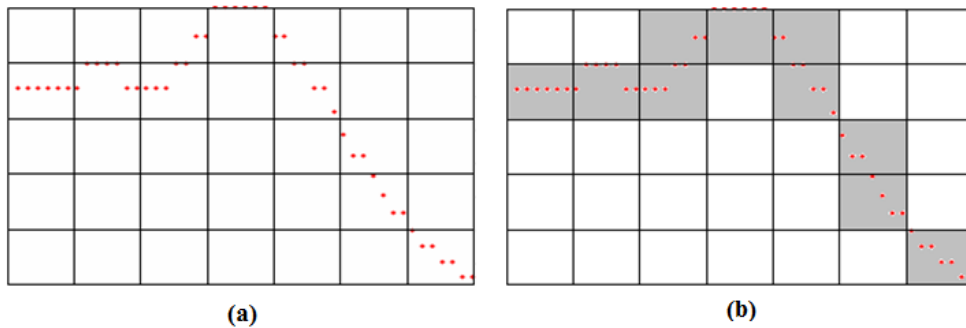
3.4. Pemodelan Jaringan Saraf Tiruan

Pemrograman jaringan saraf tiruan pada penelitian ini menggunakan bahasa pemrograman Octave. Jaringan yang dibangun akan dicoba menggunakan input berupa nilai *pitch* suara. Melalui pemrograman, akan dicari metode pelatihan (*training*) yang paling tepat dalam pengenalan jaringan

terhadap pola-pola input. Pencarian jumlah neuron lapisan tersembunyi juga akan dilakukan, agar *error* yang dihasilkan output seminimal mungkin. Data yang membangun jaringan saraf tiruan dapat dibedakan menjadi dua, yaitu data pelatihan dan data uji. Kedua jenis data ini yang akan menentukan kemampuan jaringan untuk mengenali pola intonasi yang menjadi fokus penelitian ini.

Nilai-nilai *pitch* yang diperoleh pada tiap-tiap sampel tidak secara langsung dimasukkan untuk data pelatihan. Hal ini disebabkan karena setiap sampel menghasilkan nilai *pitch* dengan jumlah yang berbeda-beda. Hal ini menyulitkan, karena yang diperlukan adalah jumlah nilai *pitch* yang sama untuk setiap data pelatihan JST. Oleh karena itu, diperlukan suatu metode yang tepat agar data memiliki jumlah yang sama, namun tetap memperlihatkan ciri dari masing-masing intonasi.

Setiap matriks input yang menjadi masukan JST dalam bahasa pemrograman Octave harus mempunyai ukuran kolom yang sama. Oleh karena jumlah nilai *pitch* yang dihasilkan berbeda pada setiap sampel suara, maka harus dipertimbangkan cara yang tepat agar JST yang dibangun mampu membedakan intonasi. Pada penelitian ini, digunakan metode pemetaan kontur *pitch* ke dalam suatu matriks. Metode ini merupakan salah satu metode yang digunakan dalam sistem pengenalan karakter (*character recognition*). Setiap barisan *pitch* yang dihasilkan akan seolah-olah dianggap sebagai suatu karakter tertentu. Berapapun ukuran rentang karakter yang dihasilkan akan dipetakan dalam suatu matriks dengan ukuran 5×7, seperti terlihat pada Gambar 5(a) dan 5(b).



Gambar 5 (a) Pemetaan deretan nilai *pitch* ke dalam matriks 5×7, (b) Penandaan kolom-kolom yang dilalui oleh deretan nilai *pitch*

Setiap kotak-kotak yang terdapat nilai *pitch* akan diberi nilai **1**, sedangkan yang lain diberi nilai **0**. Matriks yang dihasilkan akan memiliki koefisien yang khas untuk setiap intonasi yang berbeda, seperti terlihat pada Gambar 6. Koefisien matriks 5×7 yang dihasilkan selanjutnya akan diubah menjadi matriks 1 kolom. Data inilah yang akan menjadi input dalam JST.

0	0	1	1	1	0	0
1	1	1	0	1	0	0
0	0	0	0	0	1	0
0	0	0	0	0	1	0
0	0	0	0	0	0	1

Gambar 6. Binerisasi nilai matriks

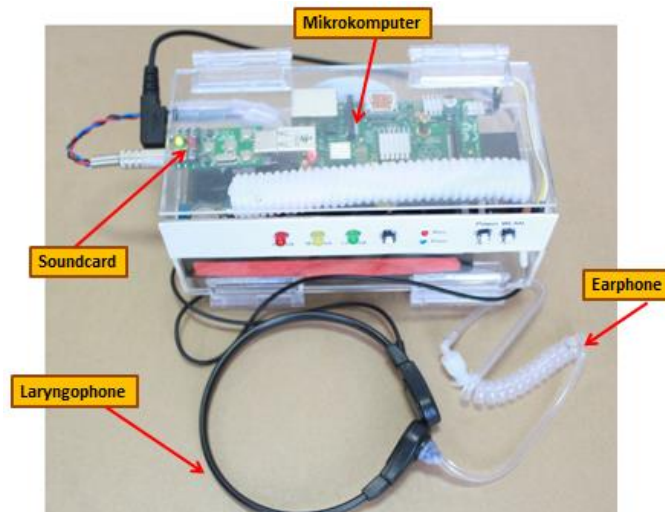
3.5. Pemodelan Jaringan Saraf Tiruan

Berdasarkan hasil pengolahan sinyal suara, terdapat dua kemungkinan respon yang diberikan oleh sistem, yaitu data input suara dinyatakan benar atau salah. Jika dinyatakan benar, maka rangkaian elektronik akan mengaktifkan relai dan selanjutnya menggerakkan solenoid *electric locker*. Dalam hal ini, sistem dimodelkan sebagai sistem akses sebuah pintu. Disisi lain, jika sistem menyatakan bahwa input suara salah, maka relai tidak akan diaktifkan.

4. Hasil dan Pembahasan

4.3 Perangkat Keras Sistem

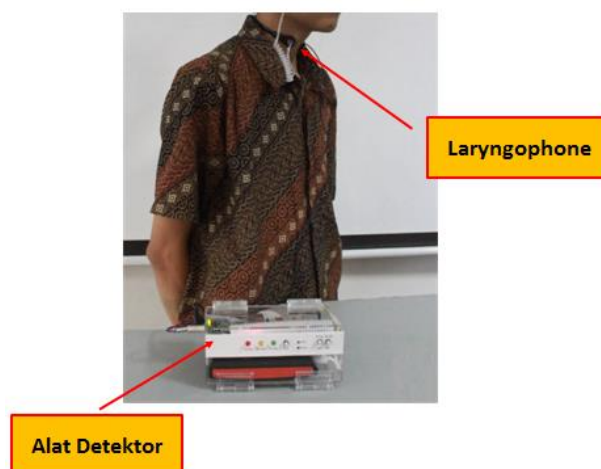
Komponen perangkat keras yang digunakan pada penelitian ini meliputi penggunaan sebuah mikrokomputer Raspberry Pi, mikrofon, *analog-to-digital converter* (ADC) dan *electric locker* beserta *driver*-nya. Setiap komponen memiliki peranannya tersendiri dan diatur sedemikian rupa dalam satu kemasan paket. Hasil perangkat keras yang tercipta diperlihatkan seperti pada Gambar 7.



Gambar 7. Perangkat Keras yang Dihasilkan

Pada sistem pengenalan suara yang memanfaatkan input berupa gelombang suara dari *user*, tantangan utama yang sering dihadapi adalah masalah derau (*noise*) dari luar. Penggunaan mikrofon genggam yang umum digunakan sangat sensitif terhadap hal ini, kecuali menggunakan mikrofon yang didesain khusus untuk hal ini. Jika menggunakan pendekatan ini, harga mikrofon sendiri akan cukup tinggi. Salah satu solusi yang diusulkan adalah dengan menggunakan mikrofon khusus, yaitu *laryngophone*. Tipe mic seperti ini digunakan pada tujuan khusus, misalnya pada pengguna radio HT (*handy talky*), sistem komunikasi untuk pilot, tentara pada medan perang, dan sebagainya. Keuntungan utama adalah mengurangi derau suara dari luar, karena alat ini akan hanya menerima getaran dada kulit leher pengguna sebagai input. Penggunaan alat ini dapat diperlihatkan pada Gambar 8.

Selanjutnya, hasil pengenalan dari alat akan dieksekusi menggunakan dua buah *electric locker*. Tidak seperti *locker* yang memanfaatkan kunci, tipe *locker* pintu jenis ini diaktifkan menggunakan sinyal listrik, sehingga otomasi sistem keamanan berbasis intonasi suara pada penelitian ini dapat diwujudkan. Pada penelitian ini, kedua pengunci diletakkan pada pintu masuk Laboratorium Hardware STMIK STIKOM Bali, seperti yang diperlihatkan pada Gambar 9.



Gambar 8. Tatacara Penggunaan Alat



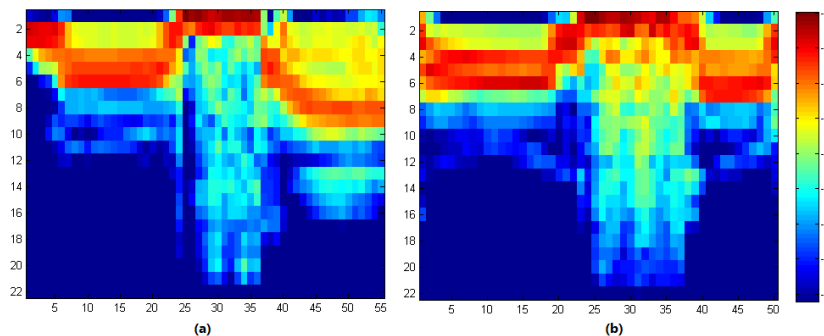
Gambar 9. Posisi Peletakkan *Electric Locker*

4.4 Desain Keamanan Berbasis Pengenalan Suara dan Intonasi

Untuk merealisasikan tujuan akhir dari penelitian ini, selain membuat perangkat keras, pada penelitian ini juga membuat perangkat lunak. Sistem keamanan berbasis intonasi ini akan menerapkan dua proses umum dalam sistem pengolahan suara, yaitu menentukan jenis kata apa yang terucap dan setelah mengetahui jenis kata yang terucap, langkah selanjutnya adalah menentukan jenis intonasi yang digunakan. Pada subbab berikut, kedua proses ini akan dijelaskan dengan lebih detail.

Matriks mel merupakan hasil terakhir dari pengolahan suara dalam proses pengenalan suara, sebelum dilakukannya proses pencarian jarak antara suara input dan suara *template*. Setiap suara input akan diubah menjadi matriks mel, dan selanjutnya dicocokkan dengan suara *template* yang juga berupa matriks mel. Proses pencocokan berupa pencarian jarak terdekat dari dua buah matriks melalui proses *dynamic programming* (DP). Matriks mel yang terbentuk merupakan sekumpulan vektor ciri. Untuk memperlihatkan matriks mel yang lebih komunikatif, maka matriks mel akan diperlihatkan secara visual, dengan melakukan penyekalaan terhadap setiap angka dalam matriks terhadap rentang angka dalam matriks tersebut. Selanjutnya, setiap skala akan dipetakan ke dalam suatu peta warna (*colormap*).

Secara umum, apabila pemotongan kata dilakukan secara tepat, maka setiap ucapan kata yang sama oleh orang yang sama akan memiliki pola matriks yang hampir sama pula. Kenyataan ini yang membuat tingkat kemiripan dari dua buah kata akan ditentukan dari jarak terpendek dua buah matriks mel dari dua kata tersebut. Pengukuran ini juga sekaligus menjadi dasar sistem *speaker dependent* (sistem tergantung pembicara). Karena kekhasan spektral suara setiap orang, maka pola pada matriks mel juga unik. Setiap spektral orang yang berbeda akan jatuh pada *filter window* yang berbeda pula, sehingga pola matriks mel setiap orang juga akan berbeda. Setiap suara input akan dilakukan proses pengolahan (*speech processing*) untuk memperoleh matriks mel-nya. Satu suara input akan dicari jarak terdekat terhadap kedua matriks mel *template* (matriks mel dari kata “buka” dan “tutup”). Jarak terdekat merupakan kata yang paling ‘mirip’, dan diharapkan sesuai dengan target yang diinginkan. Pola mel untuk kata “buka” dan “tutup” diperlihatkan pada Gambar 10.

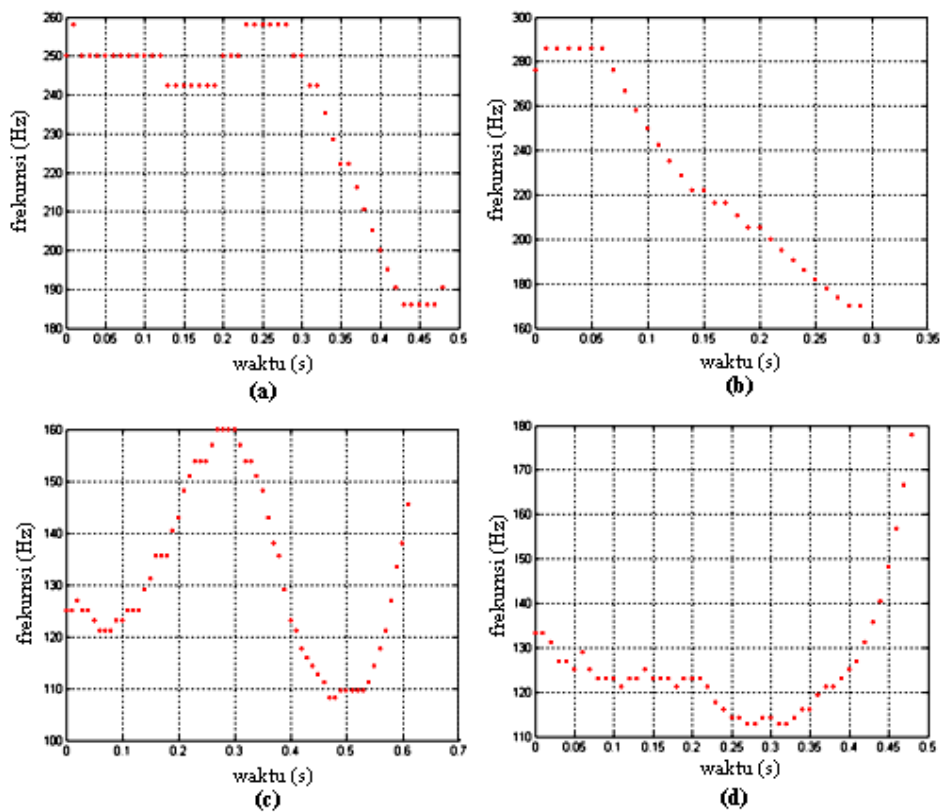


Gambar 10. Matriks Mel untuk kata (a)“buka”,(b)“tutup”

Secara umum, intonasi ucapan Bahasa Indonesia dapat dibedakan menjadi dua, yaitu intonasi berita/pernyataan (*declarative*) dan pertanyaan (*interrogative*). Setiap sampel suara yang diamati akan dicari nilai *pitch*-nya. Nilai-nilai *pitch* yang diperoleh akan diploting dalam suatu grafik dan selanjutnya dilakukan pengamatan terhadap *pitch* intonasi berita (*declarative*) dan *pitch* intonasi pertanyaan (*interrogative*). Pengamatan dilakukan untuk mencari suatu ciri yang khas dari setiap lintasan yang dibentuk oleh nilai-nilai *pitch* yang dihasilkan. Penggambaran lintasan ini dapat digunakan untuk mewakili ciri khas dari intonasi ucapan Bahasa Indonesia. Pada Gambar 11, disajikan beberapa hasil pengamatan terhadap nilai-nilai *pitch* dari dua jenis intonasi tersebut.

Diperoleh keterangan bahwa lintasan *pitch* untuk intonasi berita Bahasa Indonesia umumnya memiliki **alur menurun** pada akhir lintasan. Namun kenyataan berbeda terlihat untuk intonasi bertanya. Pada intonasi bertanya, **alur naik** justru terlihat pada akhir lintasan. Kontur nilai *pitch* ini menjadi ciri khas dari dialek Bahasa Indonesia secara umum. Tinjauan yang berbeda terhadap kontur *pitch* yang dihasilkan oleh intonasi berita dan intonasi tanya akan menjadi informasi yang penting untuk pengenalan jaringan saraf tiruan.

Fenomena naik dan turunnya kontur *pitch* pada akhir lintasan mengindikasikan pola kerja dari *vocal cord*. Secara umum, untuk intonasi berita ucapan Bahasa Indonesia, *vocal cord* menurunkan aktifitasnya vibrasinya, sehingga frekuensi fundamental juga akan menurun. Di sisi lain, ketika suatu intonasi tanya terucap, frekuensi fundamental justru akan naik pada akhir lintasan. Fenomena naik dan turunnya alur pada di akhir kata ini terlihat pada semua responden. Oleh karena itu, sulit menjadikan intonasi sebagai parameter untuk membuat suatu sistem *speaker dependent*. Pada penelitian ini, pembuatan sistem *speaker dependent* ditekankan pada pengenalan suara yang memanfaatkan metode pencarian jarak terdekat dari vektor ciri sampel uji dan vektor ciri *template*.

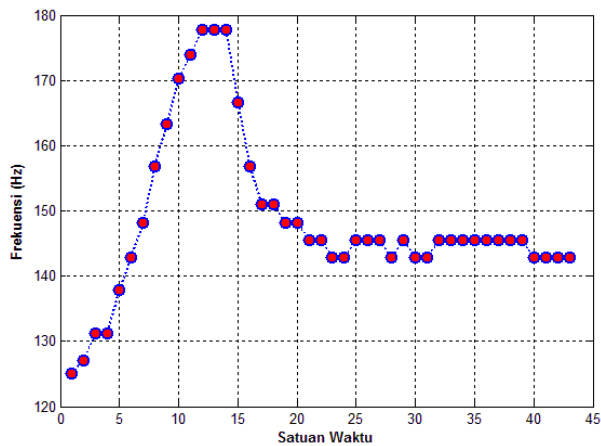


Gambar 11. Pola Pitch untuk berbagai jenis intonasi

Untuk merealisasikan tujuan penelitian, maka sistem dirancang untuk melakukan dua jenis pengenalan. Pengenalan pertama adalah pengenalan **jenis kata**, dan pengenalan kedua adalah pengenalan **intonasi**. Hal ini dilakukan sebagai filter bertingkat agar sistem keamanan yang diterapkan memiliki tingkat kesulitan yang lebih tinggi, yang dalam hal ini diperuntukkan untuk mengakses pintu Laboratorium Hardware STMIK STIKOM Bali.

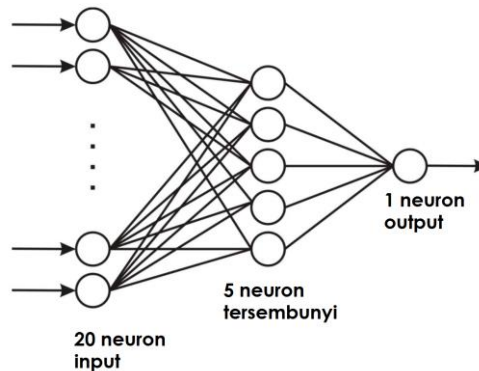
Pengenalan jenis kata berguna agar pemakai tidak dapat mengakses pintu dengan sembarang kata sandi. Kata yang dimasukkan oleh pemakai ke sistem haruslah tepat sesuai dengan apa yang telah tersimpan pada database. Pada penelitian ini, kata “buka” dan “tutup” masing-masing digunakan untuk membuka akses dan mengunci kembali pintu laboratorium. Untuk merealisasikan sistem ini, dua buah pengunci elektrik digunakan sebagai aktuator yang dengan leluasa dapat mengunci dan membuka kembali berdasarkan input sinyal elektrik yang diberikan.

Walaupun kata yang diucapkan telah benar, namun belum tentu pintu dapat diakses. Hal ini dikarenakan sistem akan mengecek jenis intonasi yang menyertai kata tersebut. Tentunya, kata “buka?” dan kata “buka!” akan memiliki intonasi yang berbeda. Pada penelitian ini, jenis intonasi yang dikenali akan bergantung pada nilai pitch yang diekstrak dari kata tersebut. Pola nilai *pitch* yang akan digunakan sebagai acuan diperlihatkan seperti pada Gambar 12. Hal ini berarti hanya intonasi-intonasi yang mendekati pola inilah yang akan dianggap benar oleh sistem.



Gambar 12. Sampel nilai *pitch* acuan

Berdasarkan hasil pengujian, maka diambil jaringan dengan **arsitektur 5-1**, kecepatan belajar (*learning rate* **0,75**) dan **momentum 0,8** sebagai data jaringan yang ditanam dalam program untuk mengenali intonasi. Arsitektur akhir jaringan terlihat pada gambar 5.10.



Gambar 5.10 Arsitektur Akhir Jaringan Saraf Tiruan

4.5 Pengujian Sistem

Pengujian akhir ditekankan pada pengenalan keseluruhan sistem, mulai dari kinerja *software* dan *hardware*. Karena sistem *hardware* yang tidak terlalu rumit, kinerja *hardware* berjalan sangat baik. Mikrokomputer dapat menjalankan tugasnya dengan baik. Pengujian justru lebih ditekankan pada pengenalan kata dan intonasi suara. Kembali uraian yang telah dipaparkan pada Bab I, penelitian ini lebih ditekankan pada sistem keamanan. Harapannya, hanya orang-orang tertentu saja yang mampu menjalankan sistem dengan suara pribadinya. Pada penelitian ini, suara penulis dijadikan sebagai pihak yang berwenang untuk menjalankan sistem. Oleh karena itu, pengenalan kemampuan mengenali suara **sebanyak mungkin** terhadap suara penulis dan pengenalan suara **sedikit mungkin** terhadap suara orang lain menjadi tinjauan keberhasilan sistem

Berdasarkan pengujian terhadap 75 percobaan memasukkan suara dari 3 orang responden, diketahui pengenalan sistem hanya **24%**, sedangkan pengujian dengan jumlah percobaan yang sama oleh penulis sendiri memiliki keakuratan sebesar **81,3%**. Hasil ini memang lebih mengarah ke *speaker dependent*, dimana sedapat mungkin sistem hanya bisa dijalankan oleh orang tertentu (dalam hal ini oleh penulis sendiri).

4. Simpulan

Terdapat beberapa kesimpulan yang dapat ditarik melalui penelitian ini, yaitu sebagai berikut:

1. Mikrokomputer dapat digunakan untuk menjalankan fungsi selayaknya komputer, walaupun dengan *resources* yang terbatas. Oleh karena itu, diperlukan program yang efisien, walaupun pengolahan suara umumnya dapat bersifat kompleks.
2. Setiap kata yang berbeda akan menghasilkan pola matriks mel yang berbeda pula. Kemiripan pola matriks mel dapat mengindikasikan kemiripan kata.
3. Jaringan Saraf Tiruan (JST) berpotensi digunakan sebagai sistem pengenalan intonasi, dan Octave merupakan perangkat lunak *opensource* yang mirip dengan MATLAB yang dapat didedikasikan untuk sistem tertanam (*embedded systems*) pada mikrokomputer.
4. Pengenalan cenderung kearah *speaker dependent* dengan tingkat keberhasilan 81,3%

5. Ucapan Terima Kasih

Peneliti menyampaikan ucapan terim kasih kepada pihak-pihak yang telah membantu pelaksanaan penelitian, baik secara finansial maupun penyediaan prasarana lainnya. Penelitian ini dibiayai oleh Direktorat Penelitian dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan dan Kebudayaan, sesuai dengan Surat Perjanjian Pelaksanaan Penugasan Penelitian Dosen Pemula Usulan Baru Bagi Dosen Perguruan Tinggi Swasta Kopertis Wilayah VIII, dengan Nomor: 1825/K8/KM/2014, tanggal 17 Juli 2014. Peneliti juga menyampaikan ucapan terima kasih kepada Ketua STMIK STIKOM Bali dan P2M STMIK STIKOM Bali karena telah menyediakan administrasi, tempat dan prasarana penelitian.

Daftar Pustaka

- [1] Lupu, E. dan Cioban, M., 2009, *Voice Biometric System*, Federation for Medical and Biological Engineering (IFMBE) Vol 26, pp. 239-242.
- [2] Hariharan, M., Yaacob, S. dan Adom, A. *A review: Malay Speech Recognition and Audio Visual Speech Recognition*. IEEE International Conference on Biomedical Engineering (ICoBE) 2012, 479-484.
- [3] Syazilawati M dan Wahyudi, 2006, *Intelligent Voice-Based Door Access Control System using Artificial Neural Network and LPC Features*. Prosiding Industrial Electronics Seminar (IES), PENS-ITS.
- [4] Wahyudi, Astuti, W., Mohamed S., 2007, *Intelligent Voice-Based Door Access Control System Using Adaptive-Network-based Fuzzy Inference Systems (ANFIS) for Building Security*, Department of Mechatronics Engineering International Islamic University Malaysia, Malaysia.
- [5] Rabiner, L., Juang, B.H., 1993, *Fundamental of Speech Recognition*, Prentice-Hall Inc., New Jersey.
- [6] Katagiri, S., 2000, *Handbook of Neural Network for Speech Processing*, Artech House Inc., Canton.
- [7] Hosom. J.P., de Villiers, J., Cole, R., Fauty, M., Schalkwyk, J., Yan, Y., Wei, W., *Training Hidden Markov Model/Artificial Neural Network (HMM/ANN) Hybrids for Automatic Speech Recognition (ASR)*. <http://www.cse.ogi.edu/class/cse552/>.
- [8] Niemann, H., Denzler, J., Kahles, B., Kompe, R., Kiessling, A., Nöth, E., Strom, V., 1994, *Pitch Determination Considering Laryngealization Effects In Spoken Dialogs*, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen.
- [9] Gerhard, D., 2003, *Pitch Extraction and Fundamental Frequency, History and Current Techniques*, Department of Computer Science University of Regina, Regina.
- [10] Wardana, I.N.K., 2008, *Identifikasi Intonasi Ucapan Bahasa Indonesia untuk Sistem Monitoring dan Pengendalian Perangkat Listrik Menggunakan Jaringan Saraf Tiruan*, Skripsi: Jurusan Teknik Fisika Universitas Gadjah Mada.
- [11] Aditya, S., Kusuma, W., *Digital Signage sebagai Media Penyampaian Informasi Kegiatan Akademik Berbasis Mikrokomputer*, Citec Journal Vol. 1 No.6, pp.308, ISSN:2354-5771.