

SISTEM KEAMANAN BERBASIS PENGENALAN SUARA SEBAGAI PENGAKSES PINTU MENGGUNAKAN METODE MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

DOOR ENTRY USING VOICE RECOGNITION SECURITY SYSTEM WHILE UTILIZING MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

Femmy Fauziah¹, Iwan Iwut Tritoasmoro², Syamsul Rizal³

^{1,2,3} Universitas Telkom, Bandung

femmyfauziah@student.telkomuniversity.ac.id¹, iwaniwut@telkomuniversity.ac.id²

syamsul@telkomuniversity.ac.id³

Abstrak

Seiring dengan perkembangan zaman serta dengan kemajuan teknologi, keamanan dapat ditingkatkan dengan memanfaatkan sinyal suara. Hal ini dilakukan karena menggunakan kunci konvensional dirasa kurang efektif. *Voice Recognition* merupakan suatu perkembangan teknik dan sistem yang memungkinkan komputer untuk menerima masukan berupa kata yang diucapkan. Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami kata-kata yang diucapkan dengan cara mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu *database*.

Dalam Tugas Akhir ini membuat suatu sistem keamanan berbasis *Voice Recognition* menggunakan metode *Mel Frequency Cepstral Coefficient* (MFCC) sebagai pengekstraksi ciri serta metode *Vector Quantization* (VQ) sebagai kuantisasi vektor. Dilakukan dengan melakukan dua tahap, yaitu tahap *training* dan tahap *testing*. Sistem buka tutup pintu secara otomatis dengan pengenalan suara menggunakan *arduino uno* dan pemrograman *python* ini memanfaatkan metode *Mel Frequency Cepstral Coefficient* (MFCC) sebagai ekstraksi ciri dan metode *Vector Quantization* (VQ) sebagai kuantisasi vektor. Pada Tugas Akhir ini menampilkan *output* MFCC dan suara *user* yang berhasil dikenali. Dilakukan pada jarak 10 cm dan 15 cm dalam keadaan hening dan ramai. Akurasi pada jarak dalam keadaan hening menghasilkan akurasi sebesar 82% dan pada keadaan ramai menghasilkan akurasi sebesar 74%.

Kata kunci: *Mel Frequency Cepstral Coefficient* (MFCC), *Vector Quantization* (VQ), *Voice Recognition*, *Python*, *Arduino*.

Abstract

Along with the times and with technological advances, security can be increased by utilizing voice signals. This is done because using conventional keys is considered less effective. *Voice Recognition* is a development of techniques and systems that allow computers to receive input in the form of spoken words. This technology allows a device to recognize and understand spoken words by matching the digital signal with a certain pattern stored in a database.

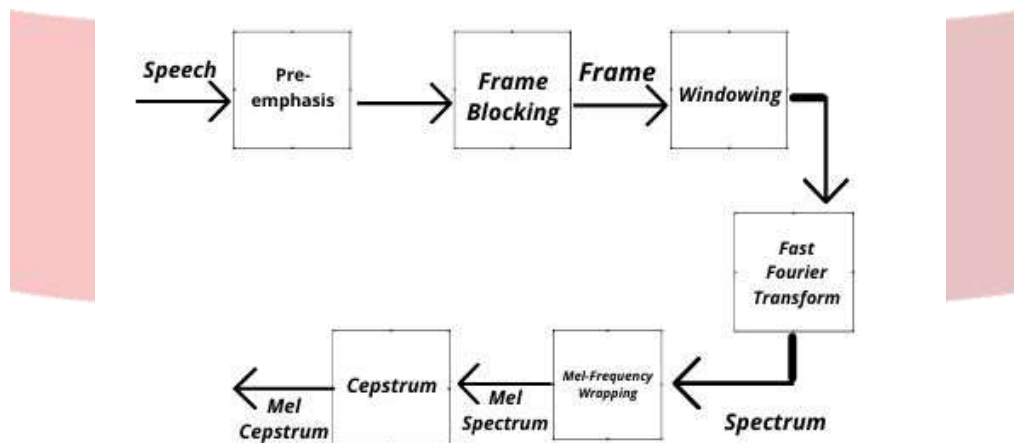
In this final project is create a security system based on *Voice Recognition* using *Mel Frequency Cepstral Coefficient* (MFCC) method as feature extraction and *Vector Quantization* (VQ) method as vector quantization. It is carried out in two stages, namely the training stage and the testing stage. This automatic door opening and closing system with voice recognition using *Arduino Uno* and *python* programming utilizes the *Mel Frequency Cepstral Coefficient* (MFCC) method as feature extraction and the *Vector Quantization* (VQ) method as vector quantization. This final project displays the MFCC output and the user's voice that has been recognized. Performed at a distance of 10 cm and 15 cm in a quiet and crowded state. Accuracy at quiet distances produces an accuracy of 82% and in crowded conditions produces an accuracy of 74%.

Keywords: *Mel Frequency Cepstral Coefficient* (MFCC), *Vector Quantization* (VQ), *Voice Recognition*, *Python*, *Arduino*.

1. Pendahuluan

Berbagai perkembangan teknologi dalam sistem keamanan pengenalan suara dapat dimanfaatkan. Salah satunya yaitu sistem pengaman pintu dengan mencocokkan karakteristik ucapan yang ada di dalam *database* dengan ucapan masukan. Teknik pengkodean yang digunakan dalam pengekstraksian ciri sinyal ucapan adalah *Mel Frequency Cepstral Coefficient* (MFCC) dan metode *Vector Quantization* (VQ) sebagai kuantisasi vektor[1].

Sistem keamanan buka tutup pintu secara otomatis dengan pengenalan suara merupakan sistem pengenalan ucapan yang digunakan sebagai otorisasi akses pintu menggunakan *Mel Frequency Cepstral Coefficient* (MFCC) dan *Vector Quantization* (VQ), sehingga hanya mengenali ucapan yang diucapkan sesuai dengan kata kunci yang telah



Gambar 2. 1 Langkah-Langkah MFCC

ditetapkan yang diharapkan mampu meningkatkan sistem keamanan serta dapat memberi manfaat dalam membantu memudahkan dalam akses buka tutup pintu.

2. Konsep Dasar

A. Mel Frequency Cepstral Coefficient (MFCC)

Mel Frequency Cepstral Coefficient (MFCC) merupakan penentu pola suara yang termasuk dalam salah satu metode ekstraksi ciri *feature extraction*. Karena sering dianggap cukup baik dalam hal mempresentasikan sinyal yang bekerja berdasarkan perbedaan pada frekuensi, metode MFCC ini paling sering dipakai pada bidang pemrosesan suara. Proses MFCC dapat dilihat pada gambar (2.1) berikut ini [2]:

A.1 Pre-emphasis

Pre-emphasis merupakan proses pengolahan sinyal suara sebelum dimodulasi yang diperlukan setelah proses sampling. Proses filter ini dilakukan untuk memperoleh spectral frekuensi sinyal menjadi lebih halus. Pada frekuensi lebih dari 2000 Hz bentuk spectral relatif turun secara tajam dan cenderung memiliki nilai tinggi pada daerah rendah. Persamaan (1) adalah rumus untuk *pre-emphasis*:

$$y(n) = x(n) - ax(n-1), 0.9 < a < 1.0. \quad (1)$$

Keterangan :

$y[n]$ = sinyal hasil *pre-emphasis*

$x[n]$ = sinyal sebelum *pre-emphasis*

a = konstanta

A.2 Frame Blocking

Pada proses *Frame Blocking*, sinyal suara dibagi kedalam beberapa *frame* yang tumpang tidih. Proses tersebut akan terus berjalan sampai seluruh sinyal masuk ke dalam frame agar tidak ada sinyal yang hilang. Hal yang

mempengaruhi hasil analisis spektral yaitu panjang *frame*, semakin panjang *frame* maka semakin baik resolusi frekuensinya. Rumus untuk menentukan *frame blocking* dapat dilihat pada persamaan (2)[3].

$$\text{Number of frames} = ((I - N) M + 1) \quad (2)$$

Keterangan :

I = sampel rate
 N = sampel poin
 $M = N/2$

A.3 Windowing

Proses *Windowing* dilakukan untuk mengurangi diskontinuitas sinyal di awal dan akhir *frame*. Proses *windowing* menggunakan *hamming window* seperti persamaan (3) :

$$w(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, \quad 0 \leq n \leq N-1 \quad (3)$$

Selanjutnya akan dilakukan proses *windowing* dengan menggunakan persamaan (4) :

$$y_1(n) = x_1(n) w(n), \quad 0 \leq n \leq N-1 \quad (4)$$

Keterangan :

$s(n)$ = nilai sampel sinyal
 N = ukuran *frame*
 $x_1(n)$ = nilai sampel *frame* sinyal 1
 $w(n)$ = fungsi *windowing*

A.4 Fast Fourier Transform (FFT)

Fast Fourier Transform (FFT) berfungsi sebagai pengubah domain waktu menjadi domain frekuensi. Proses FFT dapat dilihat dari persamaan (5) : [3]

$$F_k = \sum_{n=1}^N f(n) e^{-\frac{2\pi jkn}{N}} \quad (5)$$

Keterangan :

F_k = hasil dari proses FFT $n = 0, 1, 2, \dots, N-1$
 $k = 0, 1, 2, \dots, N-1$
 $j = -1$
 N = jumlah sampel pada masing-masing *frame*

A.5 Mel-Frequency Wrapping

Skala *mel-frequency* yaitu frekuensi rendah di bawah 1000 Hz yang bersifat linear dan frekuensi tinggi di atas 1000 Hz yang bersifat logaritmik, proses ini menggunakan *filterbank*. Persamaan (6) ini merupakan rumus untuk *Mel-Frequency Wrapping* :

$$H_i = \frac{2595 \log 10 + \frac{f_i}{700} 1}{\frac{f_i}{2}} \quad (6)$$

Keterangan :

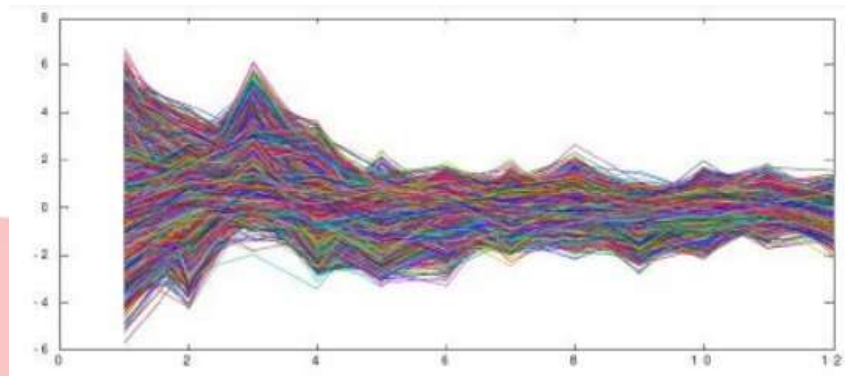
H_i = Fungsi *Mel Scale*
 f = Frekuensi

Filterbank dapat dihitung menggunakan persamaan (7) :

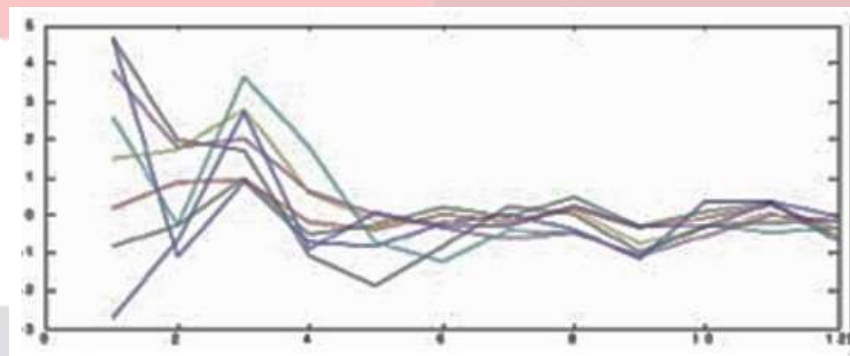
$$Y[t] = \sum_{f=1}^n S[j] H_i[j] \quad (7)$$

Keterangan :

N : jumlah besaran spektrum
 $S[j]$ = spektrum magnitudo pada frekuensi j



Gambar 3.2 Vektor sebelum mengalami proses *Vector Quantization*[5]



Gambar 3.3 Vektor sesudah mengalami proses *Vector Quantization*[5]

A..6 Cepstrum

Pada proses ini akan dikonversi *mel spectrum* menjadi domain waktu dengan menggunakan *Discrete Cosine Transform* (DCT). Fungsi dari DCT ini untuk mendapatkan nilai dimensi dari vektor yang sesuai. Hasil yang didapat dari konversi tersebut disebut *mel frekuensi cepstrum* koefisien. Rumus DCT ditunjukkan oleh persamaan (8).

$$C_n = \sum_{k=1}^K (\log S_k) \cos \left(n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right) \quad (8)$$

Keterangan :

S_k = Keluaran dari proses filterbank pada indeks k

K = jumlah koefisien

$n = 0, 1, 2, \dots, K-1$

$k = 0, 1, 2, \dots, K-1$

Hasil dari proses DCT adalah *Cepstrum* yang digunakan untuk memperoleh sinyal suara ucapan manusia serta meningkatkan kualitas proses pengenalan suara.

3. *Vector Quzationantization* (VQ)

Vector Quzationantization merupakan proses memetakan untuk mengklasifikasi vektor-vektor masukan. Tujuan dari sistem tersebut adalah untuk kompresi data. *Vector Quzationantization* akan mengekstraksi beberapa vektor corak untuk menandai karakter pembicara agar lebih spesifik[4]. Data suara pelatihan akan dikelompokkan membentuk *codebook*. Dari vektor-vektor besar menjadi terbatas yang dipresentasikan oleh pusatnya atau *codeword*. Kumpulan dari semua *codeword* tersebut adalah *codebook*. Dalam tahap pengenalan, data yang sudah didapat dari pembicara akan dibandingkan pada *codebook* untuk mengukur perbedaan dari tiap-tiap pembicara yang akan diuji. Berikut merupakan gambar perbedaan vektor sebelum mengalami proses *Vector Quantization* dan sesudah mengalami *Vector Quantization*[5]:

Algoritma LGB (*Linde, Buzo, dan Gray*) digunakan untuk menghasilkan *codebook*. Algoritma ini diimplementasikan dengan langkah-langkah sebagai berikut:

1. Rancang sebuah vektor *codebook* pertama, ini merupakan acuan dari keseluruhan masukan set pelatihan vektor ciri.
2. Gandakan ukuran *codebook* dengan membagi masing-masing *codebook* menggunakan persamaan (2.9) dan (2.10) berikut :

$$y_n^+ = y_n (1 + \epsilon) \quad (9)$$

$$y_n^- = y_n (1 - \epsilon) \quad (10)$$

Keterangan :

n = nilai antara 1 hingga M (ukuran *codebook* yang diinginkan)

$\epsilon = 0.01$

3. Temukan *codeword* terdekat untuk dijadikan titik pusat *codebook*. Untuk setiap vektor akan dikelompokkan berdasarkan *codeword* terdekat lalu akan membentuk sebuah *cluster*.
4. Pembaharuan *codeword* pada masing-masing *cluster* dengan menentukan titik pusat dari vektor ciri yang terletak di sel tersebut.
5. Iterasi 1: ulangi langkah 3 dan 4 sampai diperoleh jarak penyimpangan rata-rata yang ditetapkan.
6. Iterasi 2: ulangi langkah 2, 3 dan 4 hingga *codebook* berukuran M di desain[4].

A. Perancangan dan Realisasi Sistem

Dalam merealisasikan sistem dibutuhkan beberapa perangkat seperti dibawah ini:

1. Arduino UNO
2. Mikrofon dengan *port* USB
3. Kunci solenoid
4. Motor servo
5. Relay

Pada Gambar 2 menampilkan fase pelatihan yang dimulai dengan proses input suara individu lalu akan dilakukan proses *pre-processing*, setelah itu akan di ekstraksi ciri menggunakan MFCC untuk mengambil informasi ciri suara individu dan akan disimpan di dalam *database*.

Pada Gambar (3.4) menampilkan fase pengujian yang dimulai dengan proses input suara individu lalu akan dilakukan proses *pre-processing*, setelah itu akan di ekstraksi ciri menggunakan MFCC untuk mengambil informasi ciri suara individu dan akan dibandingkan dengan hasil ekstraksi ciri yang sudah tersimpan di *database*. Selanjutnya akan dilakukan proses kuantisasi vektor dimana jika hasilnya cocok maka kunci akan terbuka lalu pintu akan terbuka dan jika tidak cocok maka kunci dan pintu akan tetap menutup.

4. Hasil dan Analisis

A. Hasil Pengujian Program Pengenalan Suara sebagai Pengakses Pintu

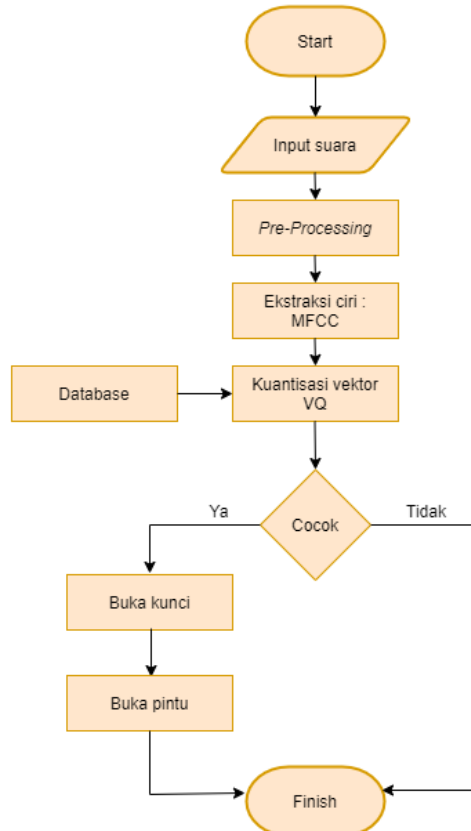
Pengujian pengenalan suara sebagai pengakses pintu ini dilakukan pada jarak 10 cm dan 15 cm dengan 2 keadaan yaitu hening dan ramai. Pengucapan kata pada setiap jarak dan keadaan dilakukan sebanyak dua puluh lima kali. Pada pengujian ini akan membandingkan antara rekaman secara langsung pengguna dengan suara yang sudah tersimpan di dalam *database*.

A.1 Pengujian pada Jarak 10 cm

Pengujian pada jarak 10 cm ini dilakukan menggunakan suara rekaman secara langsung dan ucapan suara yang telah disimpan di dalam *database* sebagai data uji. Pengujian ini dilakukan sebanyak dua puluh lima kali pada keadaan hening dan ramai. Hasil akurasi yang didapatkan oleh pengujian pada jarak 10 cm dalam keadaan hening sebesar 88% dan dalam keadaan ramai sebesar 84%. Hasil pengujian dapat dilihat pada tabel 1 dan 2.



Gambar 3. 4 Diagram Alir Fase Pelatihan



Gambar 3. 5 Diagram Alir Fase Pengujian

Tabel 4. 1 Hasil Pengujian Suara Pada Jarak 10 Cm Hening

Suara	Hasil
1	Valid
2	Valid
3	Tidak valid
4	Valid
5	Valid
6	Tidak valid
7	Valid
8	Valid
9	Valid
10	Valid
11	Valid
12	Valid
13	Valid
14	Valid
15	Tidak Valid
16	Valid
17	Valid
18	Valid
19	Valid
20	Valid
21	Valid
22	Valid
23	Valid
24	Valid
25	Valid

A..2 Pengujian pada Jarak 15 cm

Pengujian program pada jarak 15 cm dilakukan sebanyak dua puluh lima kali pengucapan dalam keadaan hening dan ramai dengan membandingkan suara rekaman langsung dengan suara yang ada di *database*. Hasil akurasi yang didapatkan pada jarak 15 cm dalam keadaan hening sebesar 76% dan dalam keadaan ramai sebesar 68%. Hasil pengujian dapat dilihat pada tabel 3 dan 4.

B. Analisis

Pada program pengenalan suara sebagai pengakses pintu ini dilakukan analisis berupa pelatihan dan pengujian.

B..1 Analisis Pelatihan Data

Pada pelatihan program ini dilakukan oleh 1 orang *user* dengan merekam kata "Zia" sesuai dengan *username* yang telah didaftarkan sebanyak tiga kali rekaman. Pelatihan suara dilakukan pada keadaan hening dan ramai dengan masing-masing jarak 10 cm dan 15 cm. Data diambil melalui rekaman yang telah disimpan dalam bentuk file wav. Setelah itu akan dilakukan ekstraksi ciri menggunakan metode MFCC dan akan disimpan di dalam *database*.

B..2 Analisa Pengujian Data

Program pengenalan suara pada pintu ini menggunakan *rate audio* sebesar 44100 Hz dan duarsi pengucapan untuk merekam suara sebesar 3 detik. Pengujian dilakukan oleh 1 *user* pada dua keadaan yaitu hening dan ramai yang masing-masing dilakukan pada jarak 10 cm dan 15 cm. Dimulai dengan ucapan individu yang direkam dengan menggunakan mikrofon, lalu sinyal suara akan diolah ke laptop dengan menggunakan perangkat lunak *python*. Setelah itu akan diteruskan kedalam mikrokontroler arduino secara serial dengan menggunakan *python* agar sinyal suara dapat diproses menjadi pengenalan suara untuk membuka pintu secara otomatis. Hasil dari pengujian pada sistem dalam keadaan hening menghasilkan akurasi sebesar 82%, sedangkan pada keadaan ramai menghasilkan akurasi sebesar 74%.

Tabel 4. 2 Hasil Pengujian Suara Pada Jarak 10 Cm Ramai

Suara	Hasil
1	Valid
2	Valid
3	Valid
4	Valid
5	Valid
6	Tidak valid
7	Tidak valid
8	Tidak valid
9	Valid
10	Valid
11	Valid
12	Valid
13	Valid
14	Valid
15	Valid
16	Valid
17	Valid
18	Valid
19	Valid
20	Valid
21	Valid
22	Valid
23	Tidak valid
24	Valid
25	Valid

5. Kesimpulan

Perancangan buka tutup pintu berbasis pengenalan suara sebagai pengakses pintu berhasil direalisasikan. Pada saat pengucapan suara harus jelas, sangat penting karena untuk berhasil menjalankan sistem dibutuhkan volume yang cukup serta suara yang jelas. Berdasarkan parameter yang digunakan dan dilihat berdasarkan jarak serta keadaan maka dapat disimpulkan bahwa semakin dekat jarak pengucapan pada mikrofon dan semakin jelas suara yang ditangkap oleh mikrofon maka akurasi yang didapat akan semakin besar. Sebaliknya, semakin jauh jarak pengucapan dan semakin tidak jelas suara yang ditangkap oleh mikrofon maka akurasi yang didapat akan semakin kecil.

Pada keadaan hening menghasilkan akurasi sebesar 82% dan keadaan ramai menghasilkan akurasi sebesar 74%. Pengujian pada jarak 10 cm dalam keadaan hening menghasilkan akurasi sebesar 88% dan keadaan ramai menghasilkan akurasi sebesar 84%. Pengujian pada jarak 15 cm dalam keadaan hening menghasilkan akurasi sebesar 76% dan keadaan ramai menghasilkan akurasi sebesar 68%.

Referensi

- [1] M. A. Ridwansyah, A. Rizal, and S. Hadiyoso, "Rancang bangun kunci berbasis suara pada pintu pintar dengan menggunakan metode mel frequency cepstral coefficient (mfcc) dan k-nearest neighbor (k-nn)," *eProceedings of Engineering*, vol. 5, no. 3, 2018.
- [2] M. Sanjaya and Z. Salleh, "Implementasi pengenalan pola suara menggunakan mel-frequency cepstrum coefficients (mfcc) dan adaptive neuro-fuzzy inference system (anfis) sebagai kontrol lampu otomatis," *ALHAZEN Journal of Physics*, vol. 1, no. 1, pp. 43–54, 2014.
- [3] Y. I. Nurhasanah, I. A. Dewi, and B. A. Saputro, "Iqro reading learning system through speech recognition using mel frequency cepstral coefficient (mfcc) and vector quantization (vq) method," *IJAIT (International Journal of Applied Information Technology)*, vol. 2, no. 01, pp. 29–42, 2018.
- [4] H. Kekre, A. A. Athawale, and G. Sharma, "Speech recognition using vector quantization," in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, 2011, pp. 400–403.

Tabel 4. 3 Hasil Pengujian Suara Pada Jarak 15 cm Hening

Suara	Hasil
1	Valid
2	Valid
3	Tidak valid
4	Valid
5	Valid
6	Valid
7	Tidak Valid
8	Tidak valid
9	Tidak valid
10	Valid
11	Valid
12	Valid
13	Valid
14	Valid
15	Valid
16	Valid
17	Valid
18	Valid
19	Valid
20	Valid
21	Valid
22	Valid
23	Valid
24	Tidak valid
25	Tidak valid

- [5] W. Setiawan *et al.*, “Perbandingan metode hidden markov model dan vector quantization untuk aplikasi identifikasi suara,” *Jurnal SPEKTRUM*, vol. 3, no. 2, pp. 107–112.

Tabel 4. 4 Hasil Pengujian Suara Pada Jarak 15 Cm Ramai

Suara	Hasil
1	Valid
2	Valid
3	Tidak valid
4	Tidak valid
5	Tidak valid
6	Valid
7	Valid
8	Tidak valid
9	Valid
10	Valid
11	Valid
12	Valid
13	Valid
14	Valid
15	Valid
16	Tidak valid
17	Valid
18	Valid
19	Valid
20	Valid
21	Valid
22	Tidak valid
23	Tidak valid
24	Valid
25	Tidak valid