

Sistem deteksi ketepatan pembacaan surah al-Kautsar berbasis kata menggunakan mel frequency cepstrum coefficient dan cosine similarity

Recites fidelity detection system of al-Kautsar verse based on words using mel frequency cepstrum coefficients and cosine similarity

Jans Hendry¹⁾, Aditya Rachman²⁾, Dodi Zulherman^{2*)}

¹⁾Program Studi Teknik Elektro, FTTE, Institut Teknologi Telkom Purwokerto
Jl. D. I. Panjaitan No. 128, Purwokerto, Indonesia 53147

²⁾Program Studi Teknik Telekomunikasi, FTTE, Institut Teknologi Telkom Purwokerto
Jl. D. I. Panjaitan No. 128, Purwokerto, Indonesia 53147

Cara sitasi: J. Hendry, A. Rachman, and D. Zulherman, "Sistem deteksi ketepatan pembacaan surah al-Kautsar berbasis kata menggunakan mel frequency cepstrum coefficient dan cosine similarity," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 1, pp. 27-35, 2020. doi: [10.14710/jtsiskom.8.1.2020.27-35](https://doi.org/10.14710/jtsiskom.8.1.2020.27-35), [Online].

Abstract - In this study, a system has been developed to help detect the accuracy of the reading of the Koran in the Surah Al-Kautsar based on the accuracy of the number and pronunciation of words in one complete surah. This system is very dependent on the accuracy of word segmentation based on envelope signals. The feature extraction method used was Mel Frequency Cepstrum Coefficients (MFCC), while the Cosine Similarity method was used to detect the accuracy of the reading. From 60 data, 30 data were used for training, while the rest were for testing. From each of the 30 training and test data, 15 data were correct readings, and 15 other data were incorrect readings. System accuracy was measured by word-for-word recognition, which results in 100 % of recall and 98.96 % of precision for the training word data, and 100 % of recall and 99.65 % of precision for the test word data. For the overall reading of the surah, there were 15 correct readings and 14 incorrect readings that were recognized correctly.

Keywords – speech signal recognition; Al-Kautsar recite; Mel coefficients; cosine similarity; MFCC

Abstrak - Dalam penelitian ini telah dibuat sistem bantu untuk mendeteksi ketepatan pembacaan Al-Qur'an pada surah Al-Kautsar berdasarkan ketepatan jumlah dan pelafalan kata dalam satu surah lengkap. Sistem ini sangat tergantung dengan ketepatan segmentasi kata yang berdasarkan envelope sinyal. Metode ekstraksi ciri yang digunakan adalah Mel Frequency Cepstrum Coefficients (MFCC), sedangkan metode Cosine Similarity digunakan untuk deteksi ketepatan pembacaannya. Dari 60 data, 30 data digunakan untuk pelatihan, sedangkan sisanya untuk pengujian. Dari masing-masing 30 data latihan

dan uji, 15 data adalah pembacaan benar dan 15 data lain adalah pembacaan salah. Akurasi sistem diukur atas pengenalan kata demi kata, yaitu menghasilkan recall dan presisi 100 % dan 98,96 % untuk data kata latihan serta 100 % dan 99,65 % untuk data kata uji. Dalam hal bacaan keseluruhan surah, ada 15 bacaan benar dan 14 bacaan salah yang dikenali dengan tepat.

Kata kunci – pengenalan sinyal wicara; pembacaan Al-Kautsar; koefisien Mel; cosine similarity; MFCC

I. PENDAHULUAN

Al-Qur'an merupakan kitab suci umat Islam, dimana di dalamnya terdapat ajaran, perintah dan larangan dalam beribadah maupun menjalani kehidupan. Tiap ayat dalam kitab suci Al-Qur'an harus dibaca dengan cara yang tepat dan benar agar maknanya tersampaikan dengan tepat, selain menghasilkan alunan yang bagus dan indah. Aspek ketepatan pembacaan Al-Qur'an yang perlu diperhatikan agar arti dan makna yang disampaikan tidak keliru adalah pelafalan, tajwid, dan panjang pendeknya bacaan. Ketidaktepatan cara membaca berdasarkan ketiga aspek tersebut dapat menyebabkan kesalahan dalam pemaknaan tiap ayat dalam kitab suci Al-Qur'an. Dalam pembelajaran pembacaan ayat-ayat Al-Qur'an, diperlukan suatu alat bantu (sistem) yang dapat mendeteksi apakah pembacaan Al-Qur'annya sudah benar atau salah [1].

Sebuah sistem klasifikasi atau pengenal pembaca (reciter) ayat Al-Quran secara otomatis telah dikembangkan dalam [2]. Sistem tersebut dirancang untuk mengenali subyek yang sedang membaca ayat Al-Quran. Metode ekstraksi ciri (*feature extraction*) yang digunakan adalah MFCC, sedangkan metode klasifikasinya adalah SVM dengan variasi parameter kernel dan *hyperparameter*. Sistem tersebut menggunakan 2135 rekaman suara dari 7 *reciter*

^{*)} Penulis korespondensi (Dodi Zulherman)
Email: zulherman.dodi@ittelkom-pwt.ac.id

terkenal dari Arab Saudi dengan pembagian data latih dan uji yakni 80 % dan 20 %. Sistem dapat mengenal *reciter* dengan akurasi 90 % untuk data uji dan 40 % - 50 % untuk data lain di luar data uji.

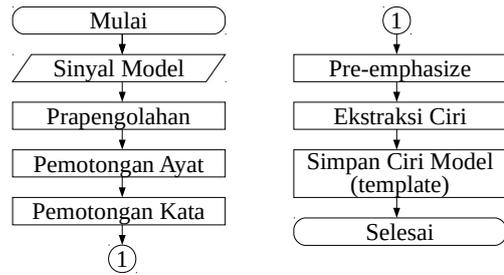
Koefisien Mel umumnya digunakan untuk ekstraksi ciri sinyal wicara (*speech signal*). Metode ini efektif untuk ekstraksi ciri dari suara pembacaan (*recite*) ayat-ayat dalam Al-Quran [2], [3]. Pengembangan sistem deteksi ketepatan pembacaan ayat-ayat dalam Quran yang bekerja secara otomatis berdasarkan pengenalan pola (*pattern recognition*) masih sangat dimungkinkan, bahkan masih menjadi fokus penelitian yang terbuka lebar [4].

Kajian terdahulu belum menghadirkan sistem otomatis yang secara lengkap dan sistematis untuk mengenali ketepatan pembacaan ayat-ayat dalam Al-Quran. Penelitian ini bertujuan mengembangkan sistem deteksi otomatis ketepatan pembacaan ayat Al-Quran berdasarkan ekstraksi ciri dari setiap kata. Surah yang dipilih adalah surah Al-Kautsar yang memiliki 3 kalimat dengan jumlah total kata sebanyak 10 dan durasi rekaman bacaan sekitar 25 - 35 detik. Metode ekstraksi ciri yang digunakan adalah MFCC dengan pemodelan berdasarkan perbedaan frekuensi yang direkam oleh telinga manusia sehingga mampu merepresentasikan ciri sinyal wicara seperti representasi manusia [5]-[8], sedangkan pengklasifikasinya menggunakan metode *cosine similarity* yang dapat mengukur jarak antara dua vektor berdasarkan besar sudut kedua vektor, bukan berdasarkan magnitudo vektor seperti dalam [9], [10]. Kelebihan dari sistem dalam penelitian ini adalah hanya menggunakan satu rekaman pembacaan oleh *expert* (ustadz) berjenis kelamin pria.

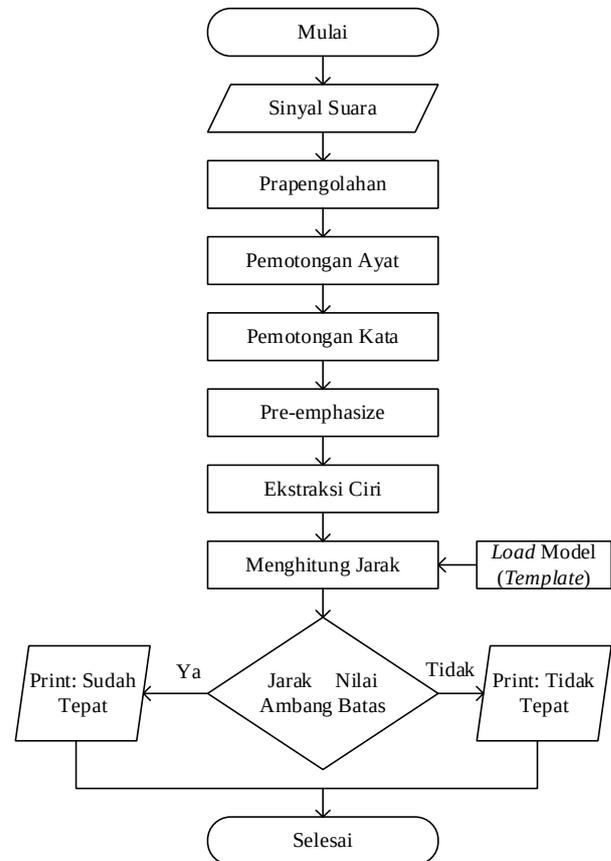
Keberhasilan sistem ini sangat dipengaruhi oleh kinerja masing-masing tahapan yang merupakan bagian dari *speech signal processing* (pengolahan sinyal wicara). Untuk itu, pemilihan metode yang tepat untuk tiap langkah tersebut sangat diperlukan. Metode yang sederhana tapi akurat adalah pilihan dalam kajian ini sehingga diperoleh sistematika metode dengan akurasi yang tinggi dan waktu proses yang cepat. Di dalam penelitian ini, ketepatan diuji berdasarkan ciri (*feature*) masing-masing kata dari surah Al-Kautsar dimana kata-kata tersebut didapatkan dari hasil segmentasi dengan melakukan teknik pemotongan kata demi kata dalam sebuah kalimat atau frasa. Metode segmentasi yang digunakan adalah *envelope detection* yang juga bertujuan untuk menghilangkan efek gema (*reverberasi*) dan derau (*noise*).

II. METODE PENELITIAN

Penelitian ini menggunakan perangkat lunak Matlab sebagai alat bantu simulasi dan sebuah laptop. Data yang digunakan dibagi menjadi dua, yaitu data latih dan data uji. Seluruh data merupakan rekaman dari pembacaan surah Al-Kautsar. Sinyal *template* berupa sebuah rekaman dari seorang *expert* (ustadz), sedangkan data latih dan uji berupa rekaman dari mahasiswa Institut Teknologi Telkom Purwokerto.



Gambar 1. Ekstraksi ciri sinyal *template*



Gambar 2. Proses pengujian sistem

Rancangan sistem ditunjukkan oleh Gambar 1 yang menunjukkan langkah-langkah pembuatan ciri *template* dan Gambar 2 yang menunjukkan diagram alir pengujian sistem. Sinyal *template* yang digunakan hanya satu sinyal bertipe *.wav yang direkam dari seorang *expert* yang juga seorang ulama. Proses perekaman dilakukan dalam kondisi hampir tidak ada derau (*noise*). Pada dasarnya sinyal wicara terbagi atas 3 bagian, yaitu *silence*, *unvoiced*, dan *voiced*. Sinyal *silence* merupakan sinyal ketika tidak ada wicara yang dihasilkan atau daya sinyalnya nol. Sinyal *unvoiced* merupakan sinyal ketika pita wicara tidak bergetar sehingga menghasilkan bentuk gelombang wicara periodik atau aperiodik yang berdaya rendah. Sinyal *voiced* merupakan bagian sinyal yang memiliki daya lebih besar dari kedua bagian tersebut [3].

A. Pengumpulan dan pengelompokan data

Perekaman penutur dilakukan secara *real time* menggunakan mikropon dengan frekuensi cuplik sebesar 44.100 Hz dan berformat *.wav. Data tutur menggunakan 10 identitas penutur berbeda dengan masing-masing membaca surah Al-Kautsar, baik laki-laki maupun perempuan, yang telah fasih membaca Al-Qur'an maupun belum. Setiap penutur direkam sebanyak 6 kali perekaman, 3 perekaman pembacaan secara benar, dan 3 perekaman pembacaan yang secara sengaja disalahkan hanya pada salah satu kata dalam surah Al-Kautsar sehingga total dari data rekaman terdapat 60 data pembacaan dengan 30 bacaan benar dan 30 bacaan salah.

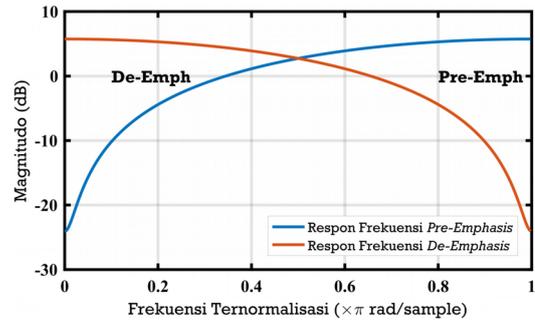
Dari 60 data, dibagi menjadi 2 kelompok, yaitu data pelatihan dan data pengujian. Terdapat 15 data bacaan benar dan 15 data bacaan salah yang dipilih secara acak digunakan untuk data pelatihan, sementara sisanya digunakan sebagai data pengujian. Dengan demikian, terdapat 30 data campuran benar dan salah untuk pelatihan dan 30 data campuran benar dan salah lainnya untuk pengujian. Pelatihan yang dimaksud dalam penelitian ini adalah langkah eksperimen dalam menentukan nilai-nilai parameter yang digunakan di dalam sistem untuk mendapatkan hasil yang optimal.

B. Prapengolahan

Bagian ini adalah tahap pengkondisian sinyal. Seluruh sinyal terlebih dahulu melalui tahap ini yang terdiri atas 3 proses, yaitu *centering*, normalisasi, dan *end-point detection* (EPD). *Centering* bertujuan untuk membuang komponen *offset* atau bias sehingga baseline seluruh sinyal akan berada pada sumbu x [11]. Normalisasi membagi seluruh sinyal dengan nilai maksimum mutlak (*absolute*) sehingga sinyal yang diperoleh berada pada range -1 Volt sampai +1 Volt seperti dalam [12]. EPD memotong sinyal wicara yang diawali dengan komponen *voiced* paling awal hingga komponen *voiced* paling akhir dari sebuah sinyal utuh. Metode yang digunakan adalah spektrogram dengan nilai ambang batas seperti dalam [13] dan *zero crossing rate* (ZCR) seperti dalam [14]. Walaupun sinyal yang direkam sudah memenuhi kriteria *centering* dan normalisasi, namun untuk mengantisipasi perbedaan spesifikasi alat perekam, maka kedua langkah tersebut tetap diimplementasikan pada sistem ini.

C. Pemotongan ayat dan kata

Surah Al-Kautsar terbagi atas 3 ayat dimana masing-masing ayat memiliki jumlah kata yang berbeda. Ayat pertama terdiri atas 3 kata, ayat kedua terdiri dari 3 kata, dan ayat terakhir tersusun atas 4 kata. Ketiga ayat disegmentasi terlebih dahulu dan tiap kata dalam tiap ayat juga disegmentasi. Segmentasi yang dimaksud adalah ekstraksi (pemotongan) ayat dan kata. Segmentasi ini dilakukan dengan menggunakan metode *envelope*. Metode yang dipilih bekerja dengan cara mencari nilai maksimum dalam panjang frame tertentu. Nilai tersebut



Gambar 3. Tanggapan frekuensi filter *pre-emphasize*

dipilih menjadi titik-titik pembentuk *envelope*. Lebar frame yang digunakan adalah 1501 titik, sedangkan nilai ambang batasnya adalah 0,01. Kedua nilai tersebut didapatkan berdasarkan eksperimen.

D. Pre-emphasize

Proses ini pada dasarnya merupakan *high pass filter* (HPF) digital yang tujuannya untuk memperbesar magnitudo sinyal pada frekuensi tinggi. Suara manusia lebih dominan pada frekuensi rendah sehingga derau menjadi dominan pada frekuensi tinggi. Untuk mengurangi pengaruh derau, maka terlebih dahulu energi sinyal pada frekuensi tinggi diperkuat menggunakan filter ini sebelum mengalami proses pengolahan lebih lanjut. Filter FIR ini diekspresikan dengan Persamaan 1 [15]. Parameter $s(n)$ dan $s'(n)$ adalah sinyal sebelum dan sesudah *pre-emphasize*. Proses ini dikenai ke setiap kata yang berhasil dipotong. Nilai yang digunakan adalah 15/16 atau sekitar 0,9375. Filter ini biasanya diikuti dengan filter *de-emphasize* yang berfungsi kebalikan dari filter *pre-emphasize*.

$$s'(n) = s(n) - \alpha \times s(n-1) \quad (1)$$

Kedua filter tersebut memiliki tanggapan frekuensi seperti yang ditunjukkan oleh Gambar 3. Tanggapan frekuensi filter *pre-emphasize* menunjukkan bahwa ini adalah tanggapan filter berjenis *high pass filter*. Dalam penelitian ini, filter *de-emphasize* tidak digunakan karena fungsinya adalah merekonstruksi sinyal yang terkena filter.

E. Ekstraksi Ciri

Ekstraksi ciri dikenai pada tiap-tiap kata yang berhasil disegmentasi. Metode ekstraksi ciri yang digunakan adalah MFCC. Metode ini memberikan ciri berupa larik berisi koefisien *cepstrum* atau koefisien Mel. Tiap kata memiliki ciri berukuran 1 x 13, namun hanya 12 elemen larik yang digunakan yakni elemen kedua hingga ketiga belas. Elemen pertama merupakan komponen *offset* atau bagian sinyal yang memiliki energi sangat besar bila dibandingkan dengan kedua belas elemen lainnya. Elemen tersebut tidak bisa dipakai karena perbedaan magnitudo yang terlalu besar dibandingkan dengan elemen lainnya yang menyebabkan kurva ciri elemen lain tampak datar jika ditampilkan pada koordinat kartesian. Langkah-

langkah untuk mendapatkan koefisien Mel adalah *framing*, *windowing*, *fast Fourier transform* (FFT), *Mel filterbank*, *log frequency*, dan *discrete cosine transform* (DCT).

Dalam *framing*, sinyal wicara dibagi menjadi banyak frame dengan setiap frame berukuran N titik dan frame tetangganya diberi jarak M titik. Ukuran frame yang dipilih tidak terlalu besar dan tidak terlalu kecil, mengikuti karakteristik dari sinyal wicara sebagai sinyal *quasi-stationary*, yaitu pada jangka waktu pendek tertentu dan sinyal masih bersifat tetap (*stationary*).

Setiap frame dibobotkan secara matematis menggunakan teknik *windowing*. Teknik *windowing* yang digunakan adalah *Hamming window* seperti dalam [7]. *Windowing* bertujuan untuk memastikan kontinuitas sinyal frame demi frame. Tipe *Hamming* menghasilkan *sidelobe* level sekitar -43 dB yang dinyatakan dalam Persamaan 2 dengan $0 \leq n \leq N-1$ [16]. Variabel w berisi koefisien *windowing*, variabel n merupakan tiap-tiap titik, sedangkan variabel N adalah jumlah titik dalam sebuah frame. Persamaan 2 menghasilkan koefisien *window* yang panjangnya $N + 1$. Bentuk kurva *Hamming Window* ditunjukkan oleh Gambar 4 dengan menggunakan nilai $N = 100$.

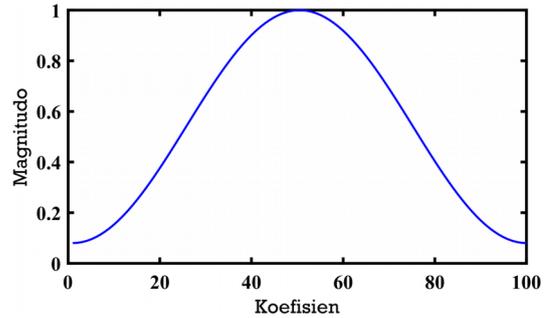
$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (2)$$

Setiap frame dianalisis menggunakan alihragam Fourier yang mengubah sinyal dari kawasan waktu ke kawasan frekuensi. Dengan demikian, ciri yang diekstrak adalah representasi kompleks sinyal untuk memudahkan proses pengenalan tutur.

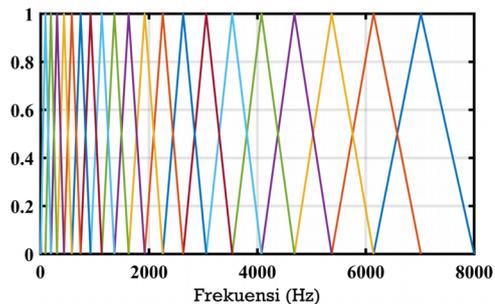
Spektrum frekuensi dari setiap *window* diubah ke dalam skala Mel yang meniru skala pendengaran manusia atau *wrapping* dengan menggunakan *triangular bandpass filterbank* seperti dalam [6]. Dengan menggunakan filter ini, besar energi pada band frekuensi tertentu sinyal wicara dapat dihitung. Filter yang digunakan berupa susunan filter segitiga yang saling tumpang tindih atau yang disebut sebagai *Mel-spectrum* seperti ditunjukkan dalam Gambar 5.

Berdasarkan hasil simulasi dalam penelitian ini, terdapat dua keuntungan ketika menggunakan *triangular filterbank*. Jumlah ciri yang digunakan menjadi lebih sedikit sehingga proses pengenalan bisa dilakukan lebih cepat. Sampul (*envelope*) dari spektrum dan harmonik juga bisa diperoleh dengan jelas. Filter ini memperhalus magnitudo spektrum sehingga komponen harmoniknya berkurang sehingga *pitch* dari tiap sinyal dihilangkan dan setiap sinyal wicara yang berbeda *pitch* tapi *timbre*-nya sama dianggap sama. Dengan kata lain, metode MFCC secara otomatis menghilangkan pengaruh *pitch* pada vektor ciri yang dihasilkannya.

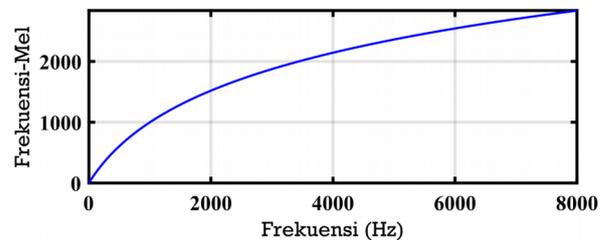
Agar perbedaan magnitudo tiap spektrum Mel tidak terlalu besar, maka dikenai dengan operasi logaritmik dengan mengambil nilai Log dari tiap-tiap magnitudo tersebut yang diekspresikan dalam Persamaan 3 seperti dalam [1]. Frekuensi Mel $mel(f)$ berbanding lurus dengan logaritma natural dari frekuensi linier (f).



Gambar 4. Bentuk kurva *Hamming window*



Gambar 5. Susunan *triangular bandpass filterbank*



Gambar 6. Hubungan antara *Mel-Cepstrum* dan frekuensi linier (Hz)

Perubahan nilai yang halus seiring dengan meningkatnya nilai frekuensi linier.

$$mel(f) = 1125 \times \ln\left(1 + \frac{f}{700}\right) \quad (3)$$

Gambar 6 menunjukkan hubungan antara spektrum frekuensi linier (Hz) dan spektrum frekuensi Mel. Bentuk kurva ini lebih disebabkan fungsi logaritma natural yang mengubah *shape* dari kurva menjadi non-linier. Ilustrasi tersebut juga menggambarkan bentuk sesungguhnya tanggapan telinga manusia terhadap perubahan frekuensi.

Nilai-nilai logaritmik frekuensi hasil proses sebelumnya didekorelasi menggunakan alihragam (transformasi) fungsi genap (kosinus) menggunakan DCT sehingga dihasilkan representasi sinyal wicara dalam bentuk bilangan riil [5]. Perhitungan ini diekspresikan dalam Persamaan 4. Parameter $C(m)$ adalah koefisien DCT, E_k adalah ${}^{20}\log$ dari hasil *triangular filterbank*, N adalah jumlah *triangular* dalam filterbanks, m memiliki nilai dari 1, 2, ... L , dan L adalah jumlah koefisien Mel yang diinginkan. Di dalam

kajian ini, jumlah *triangular* yang digunakan 20, sedangkan jumlah koefisien Melnya adalah 13.

$$C(m) = \sum_{k=1}^N E_k \cos \left[\frac{m\pi}{2N} * (2k-1) \right] \quad (4)$$

Hasil dari proses DCT adalah sebuah matriks berukuran $I \times J$ dengan I adalah baris dan J adalah kolom. Koefisien cepstrum merupakan hasil akhir dari metode MFCC sehingga disebut sebagai ciri (*features*) yang digunakan sebagai representasi baru sinyal. Di dalam kajian ini, nilai J adalah 13 sehingga dengan menghitung nilai rata-rata untuk tiap matriks *cepstrum* terhadap jumlah kolom, maka didapatkan ciri akhir dari tiap kata berupa larik berukuran 1×13 .

F. Menghitung jarak

Metode pencocokan untuk menguji ketepatan pembacaan Surah Al-Kautsar yang digunakan dalam penelitian ini adalah metode penghitungan jarak *cosine similarity*. Metode ini bekerja dengan cara menghitung tingkat kesamaan antara dua buah larik. Dalam hal ini, larik-larik yang dihitung jaraknya merupakan nilai *cepstrum* hasil MFCC yakni antara *cepstrum* sinyal *template* dan *cepstrum* sinyal masukan.

Metode *cosine similarity* dinyatakan dalam Persamaan 5. Variabel C merupakan hasil perhitungan jarak. Variabel q dan d adalah larik-larik yang akan dibandingkan untuk diukur jaraknya. Variabel i dan j melambangkan tiap-tiap elemen pada larik q dan d sedangkan t merupakan ukuran dari larik.

$$C(d_i, q_i) = 1 - \frac{q_i \times d_i}{|q_i| |d_i|} \\ = 1 - \frac{\sum_{j=1}^t (d_{ij} d_{ij})}{\sqrt{\sum_{j=1}^t (q_{ij})^2 \cdot \sum_{j=1}^t (d_{ij})^2}} \quad (5)$$

G. Recall dan Precision

Kinerja dari sistem ini diukur menggunakan parameter *recall* dan *precision*. Dalam perhitungannya, alat ukur kinerja ini melibatkan nilai-nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Parameter *recall* dan *precision* dihitung menggunakan Persamaan 6 dan Persamaan 7 [17].

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

H. Menentukan Nilai Ambang Batas

Pada proses ini, jarak antara ciri sinyal model dengan sinyal pelatihan maupun sinyal uji dilakukan. Metode yang digunakan adalah *cosine similarity* yang dihitung menggunakan Persamaan 5 yang mengukur kemiripan antara 2 buah larik. Prinsip ini digunakan untuk menentukan ketepatan pembacaan Surah Al-Kautsar kata demi kata. Untuk menentukan apakah kata tersebut tepat pembacaannya, maka sebuah nilai ambang batas (*threshold*) ditentukan berdasarkan eksperimen menggunakan kurva ROC (*Receiver Operating Characteristic*). Kurva ini dihasilkan dari nilai FPR (*False Positive Rate*) dan TPR (*True Positive Rate*) yang dapat menunjukkan letak nilai ambang terbaik untuk *precision* dan *recall* yang dinyatakan dalam Persamaan 8 dan Persamaan 9 [17]. Sebuah pembacaan dikatakan tepat atau benar jika masing-masing kata yang terdapat pada setiap ayat dikenali sebagai benar oleh sistem.

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

I. Evaluasi Kinerja Sistem

Data latih digunakan sebagai data untuk mendapatkan nilai ambang batas yang terbaik dalam menentukan ketepatan pembacaan kata demi kata. Data uji digunakan untuk menguji kinerja sistem yang sudah diberikan nilai-nilai parameter terbaik. Evaluasi kinerja dari sistem baik saat pelatihan dan pengujian dilakukan dengan menghitung *precision* dan *recall*. Kedua parameter kinerja tersebut dipilih karena pada dasarnya kasus ini merupakan kasus *binary classifier* (benar atau salah) dan melibatkan empat variabel penting dalam kasus pendeteksi/pengenal, yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

III. HASIL DAN PEMBAHASAN

Surah yang dijadikan bacaan dalam kajian ini adalah surah Al-Kautsar yang terdiri dari 3 ayat dengan total kata sejumlah 10 kata. Sinyal *template* berupa rekaman sebanyak satu rekaman dari pembacaan Surah Al-Kautsar oleh seorang *expert* berjenis kelamin pria. Sinyal untuk pelatihan dan pengujian merupakan rekaman dari 10 orang, yaitu 5 laki-laki dan 5 perempuan yang merupakan mahasiswa di Fakultas Teknik Telekomunikasi dan Elektro, Institut Teknologi Telkom Purwokerto.

Setiap individu membaca surat tersebut sebanyak 6 kali perekaman dimana 3 rekaman merupakan pembacaan yang benar, sedangkan sisanya dengan cara yang salah. Total data latih dan uji adalah 60 rekaman. Data latih digunakan untuk perancangan sistem. Dalam penelitian, data ini digunakan untuk menentukan nilai ambang batas terhadap keputusan tepat atau tidaknya sebuah bacaan. Data uji digunakan untuk mengukur

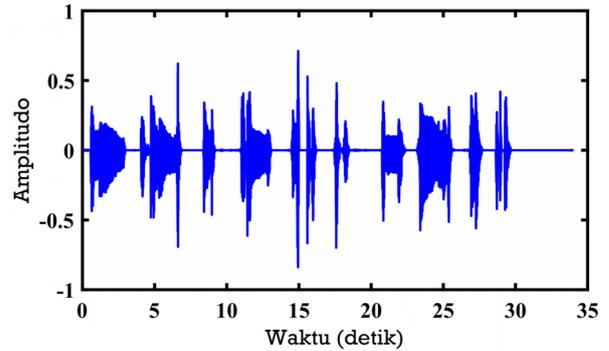
kinerja sistem secara keseluruhan setelah semua parameter terbaik digunakan.

Salah satu sinyal audio yang digunakan dalam penelitian ini dan digunakan sebagai *template* ditunjukkan oleh Gambar 7. Durasi sinyal tersebut sekitar 32-35 detik dengan nilai yang sudah dinormalisasi dan *centered*. Secara visual, tampak bahwa masih ada komponen *silence* dan *unvoiced* di ujung awal dan akhir sinyal. Bagian ini harus dihilangkan menggunakan metode EPD. Hasil dari proses ini ditunjukkan pada Gambar 8.

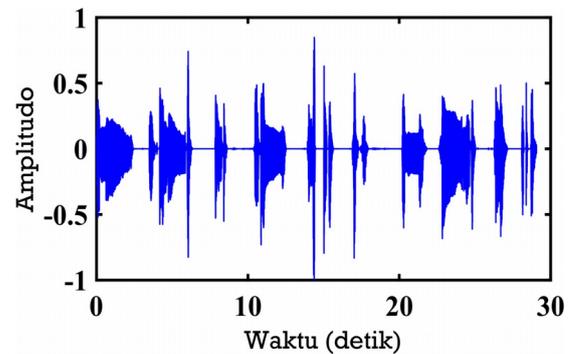
Bagian ujung dari awal dan akhir sinyal sudah terpotong yang meninggalkan hanya bagian sinyal yang dibutuhkan. Sinyal ini kemudian dipotong menjadi 3 buah ayat yang sekaligus membuang komponen *silence* dan *unvoiced*, baik antara ayat maupun di awal atau akhir Surah sehingga panjang sinyal lebih pendek daripada sinyal sebelum proses ini. Tidak ada perubahan pada amplitudo maupun frekuensi sinyal setelah melalui proses ini. Hasil dari segmentasi ayat ini dipisahkan dalam 3 warna seperti yang ditunjukkan pada Gambar 9 menjadi warna biru untuk ayat pertama, warna hijau untuk ayat kedua, dan warna merah untuk ayat ketiga.

Tiap ayat kemudian disegmentasi menjadi kata-kata penyusunnya dengan menggunakan metode yang sama dengan segmentasi ayat. Hasil segmentasi kata ditunjukkan secara berurutan pada Gambar 10. Sama halnya dengan segmentasi ayat, hasil dari segmentasi kata tidak mengalami perubahan dalam amplitudo maupun frekuensi. Perubahan terjadi pada panjang masing-masing ayat karena adanya penghilangan komponen *silence* dan *unvoiced* yang terdapat di antara masing-masing kata. Tidak adanya perubahan pada spektrum frekuensi masing-masing ayat disebabkan kecilnya energi komponen *silence* dan *unvoiced* sehingga keberadaan komponen tersebut tidak memberikan pengaruh yang signifikan, demikian pula penghilangannya.

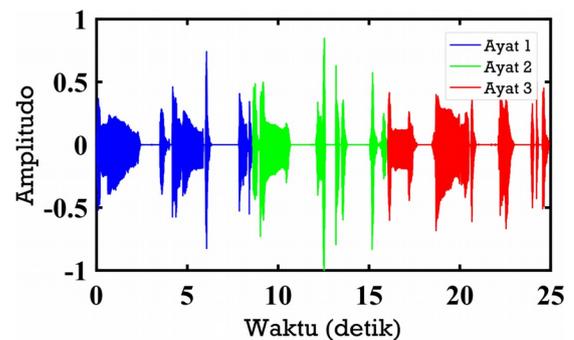
Proses berikutnya adalah *pre-emphasize* yang diberikan pada tiap-tiap kata. Dengan menggunakan kata pertama dari ayat pertama, pengaruh dari filter ini pada kawasan waktu ditunjukkan dalam Gambar 11. Amplitudo sinyal setelah proses ini menjadi semakin kecil yang menandakan bahwa distribusi sinyal wicara dominan pada frekuensi rendah sehingga ketika diberikan filter *pre-emphasize* yang juga merupakan *high pass filter*, maka sinyal pada frekuensi rendah mengalami pelemahan (*attenuation*). Di sisi lain, bagian



Gambar 7. Sinyal *template* dalam kawasan waktu



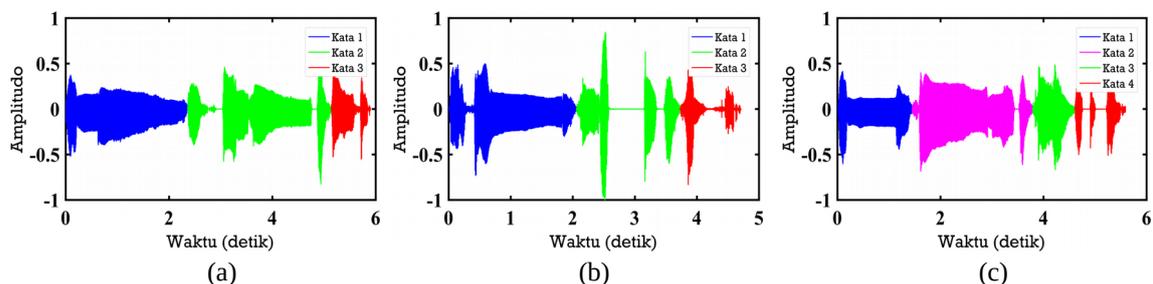
Gambar 8. Hasil pemotongan menggunakan EPD



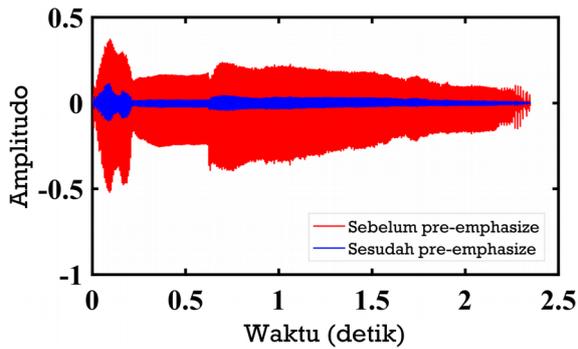
Gambar 9. Hasil segmentasi ayat

sinyal pada frekuensi tinggi diloloskan yang jika didengarkan hasilnya berbeda dengan sinyal awal.

Untuk setiap kata setelah proses *pre-emphasize*, ciri khasnya diekstrak menggunakan MFCC. Proses ini berlangsung secara berulang (*iterative*) kata demi kata. Hanya 12 koefisien *cepstrum* yang digunakan untuk merepresentasikan sebuah kata yang dimulai dari



Gambar 10. Hasil segmentasi kata: a) ayat pertama, b) ayat kedua, dan c) ayat ketiga



Gambar 11. Perbandingan sinyal sebelum dan sesudah *pre-emphasize*

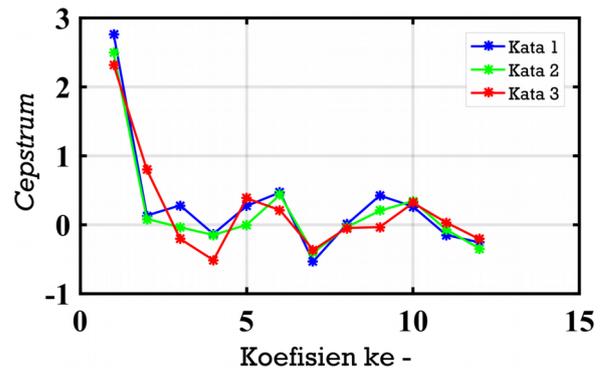
koefisien kedua hingga ketiga belas. Koefisien pertama tidak digunakan karena merupakan komponen *offset* dengan energi paling besar yang menyebabkan rasio dirinya terhadap koefisien lain sangat besar. Hal ini menyebabkan pola yang seharusnya berupa kurva fluktuatif menjadi tampak datar. Koefisien pertama ini juga menunjukkan distribusi energi sinyal dominan pada frekuensi rendah. Gambar 12 menunjukkan kurva yang terbentuk dari koefisien *cepstrum* ini.

Ciri untuk setiap kata pada tiap ayat memiliki pola yang berbeda. Dengan demikian, MFCC dinyatakan mampu menghasilkan representasi unik dari setiap kata sehingga *cepstrum* dapat digunakan sebagai materi dalam sistem ini seperti dalam [5]-[8]. Setiap koefisien pertama memiliki nilai yang cukup tinggi dibandingkan dengan koefisien lainnya. Namun, pola yang terbentuk dari tiap-tiap koefisien masih dapat dibedakan dengan menggunakan koefisien kedua dari 13 koefisien *cepstrum*. Gambar 13 berisi kurva dari 4 kata yang berasal dari ayat ketiga yang dibentuk dari 13 koefisien *cepstrum*. Koefisien pertama yang memiliki magnitudo sangat besar diperbandingkan sehingga kurva yang terbentuk oleh koefisien lain mendekati garis datar.

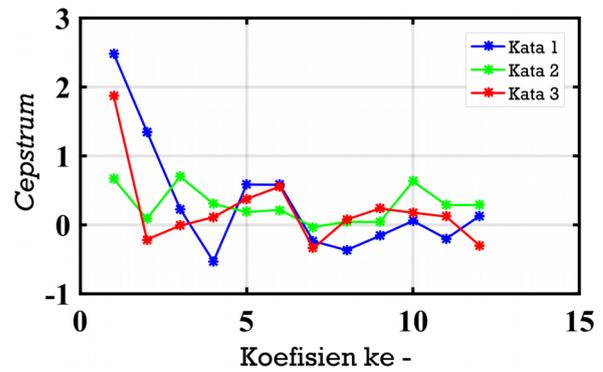
Nilai yang tampak pada grafik negatif karena telah mengalami tahapan proses MFCC. Namun, dalam kawasan frekuensi (Hz) nilai ini memiliki magnitudo spektrum paling besar. Salah satu contoh adalah kata ke empat dimana koefisien pertama bernilai 25,8871 yang sangat besar dibandingkan dengan koefisien lainnya, bahkan dengan koefisien kedua yang bernilai 1,8565 dan merupakan koefisien kedua terbesar. Hal ini berarti tidak ada pola yang jelas yang membedakan masing-masing kata dalam 1 ayat, apalagi jika dibandingkan dengan kata-kata berbeda pada ayat yang lainnya. Dari hal tersebut, koefisien *cepstrum* yang tepat sebagai ciri adalah koefisien kedua hingga ketiga belas.

A. Pelatihan

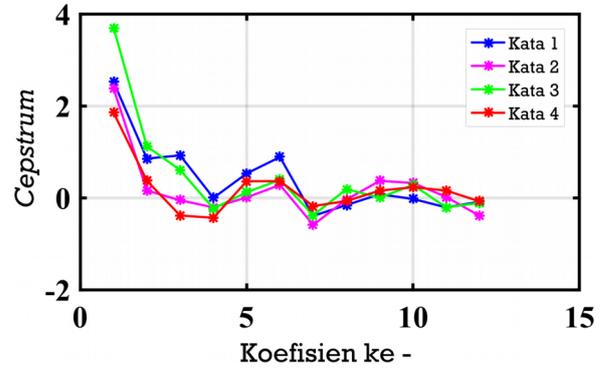
Data sejumlah 30 buah digunakan dalam tahap pelatihan ini untuk menentukan nilai ambang batas pada perhitungan jarak *cosine similarity*. Untuk memilih nilai ambang batas terbaik, disediakan pilihan nilai dalam range 0,01 hingga 1 dengan kelipatan (*increment*) sebesar 0,01. Hal ini berarti terdapat 100 nilai ambang batas yang diuji. Dari 30 data pelatihan, terdapat 15



(a)

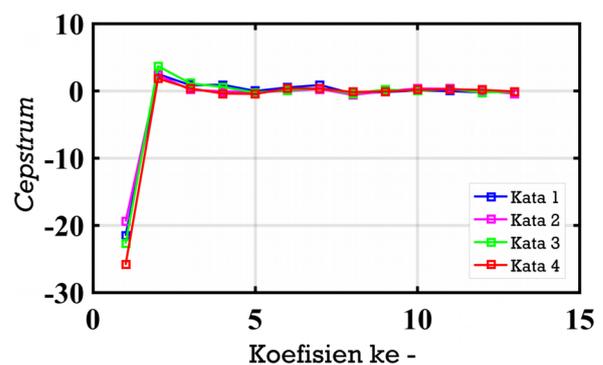


(b)



(c)

Gambar 12. Ciri dari masing-masing kata pada setiap ayat: a) ayat pertama, b) ayat kedua, dan c) ayat ketiga



Gambar 13. Koefisien pertama menyebabkan pola menjadi garis datar

bacaan benar dan 15 bacaan salah. Dari masing-masing bacaan terdapat 3 ayat yang berisi 10 kata. Untuk setiap

bacaan salah, hanya terdapat 1 kata yang disalahkan dan lokasinya acak sehingga total kata yang benar adalah 285 kata dan kata yang salah adalah 15 kata. Jumlah semua kata adalah 300 kata.

Nilai ambang batas didapat melalui observasi kurva ROC yang terbentuk oleh FPR dan TPR. Target dari observasi ini adalah nilai TPR yang besar diikuti dengan nilai FPR yang kecil. Nilai ambang batas ini bisa saja banyaknya lebih dari 1. Pada kurva ROC (**Gambar 14**), nilai ambang batas yang terbaik adalah 0,57, 0,58, 0,59 dan 0,60 karena salah satu nilai tersebut dapat dipilih untuk memberikan respon *precision* dan *recall* yang terbaik. Untuk nilai ambang lain, nilai TPR dan FPR tidak pernah bertemu pada kondisi terbaik, misalnya saat TPR tinggi, maka FPR juga tinggi, sedangkan yang dibutuhkan adalah nilai TPR yang tinggi dan FPR rendah. Dalam penelitian ini, nilai ambang batas yang dipilih adalah 0,58.

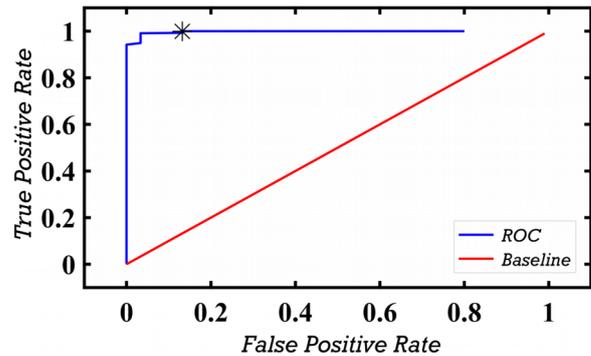
Kurva ROC ini berbentuk *imbalanced* karena nilai TP dan TN tidak sama. Nilai FP dan FN juga tidak sama. Nilai-nilai parameter tersebut dipaparkan pada **Tabel 1**. Hasil tersebut menunjukkan akurasi dalam bentuk *recall* dan *precision* dari data latih yang digunakan untuk mendapatkan nilai ambang batas yang digunakan dalam tahap pengujian sistem. Hasil pengujian menunjukkan bahwa nilai ambang batas 0,58 memberikan nilai *recall* 100 % yang berarti seluruh bacaan benar berhasil dideteksi sebagai benar. Nilai ambang batas ini ditetapkan sebagai nilai yang paling bagus untuk digunakan dalam pengujian sistem.

B. Pengujian

Data berbeda sejumlah 30 buah digunakan dalam pengujian. Dari data tersebut, terdapat 15 bacaan benar dan 15 bacaan salah. Masing-masing bacaan terdiri dari 3 ayat yang memiliki 10 kata. Sama halnya dengan pelatihan, hanya ada 1 kata yang disalahkan dan lokasinya acak sehingga total kata yang benar adalah 285 kata dan kata yang salah adalah 15 kata. Jumlah semua kata adalah 300 kata.

Berdasarkan **Tabel 2**, dengan menggunakan nilai ambang batas 0,58, terdapat 150 kata bacaan benar yang dikenali sebagai kata yang benar (TP), sedangkan dari 150 kata bacaan yang salah, sebanyak 135 kata dikenali sebagai kata yang benar (TP), 14 kata salah terdeteksi sebagai salah (TN), dan sisanya sebanyak 1 kata dideteksi sebagai kata yang benar (FP). Hal ini berarti ada 285 kata yang sudah tepat terdeteksi sebagai benar, 14 kata salah yang tepat terdeteksi sebagai salah, sedangkan 1 kata salah terdeteksi sebagai benar.

Pembacaan surah Al-Kautsar dikatakan sudah benar jika setiap kata penyusun setiap ayat pada surah tersebut dikenali sebagai bacaan yang tepat atau benar. Dengan demikian, berdasarkan hasil deteksi ketepatan pada masing-masing kata pada setiap ayat utuh bacaan surah Al-Kautsar, dari 15 data bacaan benar semuanya terdeteksi dengan baik sebagai bacaan Surah yang benar, sedangkan dari 15 bacaan salah, sebanyak 14 bacaan yang dideteksi sebagai bacaan yang salah dan 1 bacaan salah dikenali sebagai bacaan yang benar.



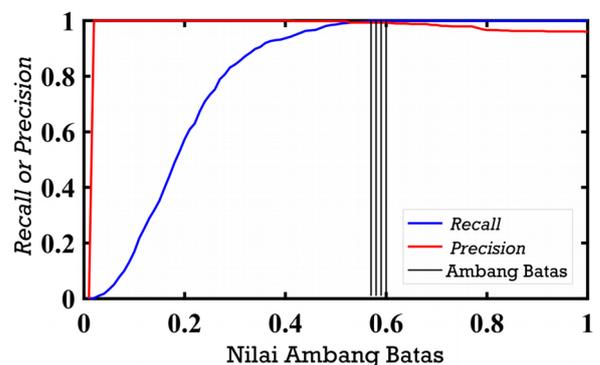
Gambar 14. Kurva ROC TPR dan FPR

Tabel 1. Hasil *recall* dan *precision* katd dari data latih

No	Parameter	Hasil
1.	TP	285
2.	TN	12
3.	FP	3
4.	FN	0
5.	<i>Recall</i> kata	100 %
6.	<i>Precision</i> kata	98,96 %

Tabel 2. Hasil *recall* dan *precision* kata dari data uji

No	Parameter	Hasil
1.	TP	285
2.	TN	14
3.	FP	1
4.	FN	0
5.	<i>Recall</i> kata	100 %
6.	<i>Precision</i> kata	99,65 %



Gambar 15. Kurva *recall* dan *precision* terhadap nilai ambang batas terbaik

Dengan kata lain, *precision* dan *recall* bacaan yang diperoleh berurutan sebesar 100 % dan 93,75 %.

Hasil ini menunjukkan bahwa nilai ambang batas yang dipilih telah memberikan hasil deteksi yang signifikan terhadap pembacaan surah Al-Kautsar. Pada **Gambar 15** ditunjukkan posisi terbaik dari tiga parameter yakni *recall*, *precision*, dan nilai ambang batas. Garis-garis vertikal hitam menandakan lokasi empat nilai ambang batas terbaik.

Nilai *precision* dan *recall* yang diperoleh pada pengujian ini menunjukkan unjukkerja sistem yang

sangat baik dalam mendeteksi ketepatan pembacaan surah Al-Kautsar jika dibandingkan [2] dengan nilai akurasi hanya 90 %. Metode MFCC mampu menghasilkan representasi unik dari setiap kata dan bacaan surah ini seperti dalam [5]-[8] dan penghitungan jarak dengan *cosine similarity* mampu menentukan nilai ambang batas seperti [9], [10]. Rancangan dalam penelitian ini dapat digunakan sebagai sistem pendeteksi kebenaran pembacaan Al-Quran secara otomatis seperti [2]-[4] dengan tingkat *recall* dan *precision* yang tinggi dan nilai akurasi lebih besar dari 90 %.

IV. KESIMPULAN

Sistem deteksi ketepatan pembacaan surat Al-Kautsar menggunakan teknik pemotongan kata, ekstraksi ciri MFCC, dan penghitungan jarak *cosine similarity* bekerja dengan sangat baik dengan nilai *recall* untuk data uji berupa 300 kata adalah 100 % dan *precision* sebesar 99,65 %. Dalam keseluruhan pembacaan, sistem mempunyai *precision* dan *recall* bacaan sebesar 100 % dan 93,75 %.

UCAPAN TERIMA KASIH

Terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM), Fakultas Teknik Telekomunikasi dan Elektro (FTTE), serta *expert* dan mahasiswa/i yang terlibat dalam pengambilan data di Institut Teknologi Telkom Purwokerto (ITTP).

DAFTAR PUSTAKA

- [1] A. Muhammad, Z. U. Qayyum, W. I. Mirza, S. Tanveer, M-. Enriquez, and A. Z. Syed, "E-hafiz: Intelligent system to help muslims in recitation and memorization of Quran," *Life Science Journal*, vol. 9, no. 1, pp. 534-541, 2012.
- [2] A. Elnagar, R. Ismail, B. Alattas, and A. Alfalasi, "Automatic classification of reciters of Quranic audio clips," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications*, Aqaba, Jordan, Nov. 2018, pp. 1-6. doi: [10.1109/AICCSA.2018.8612829](https://doi.org/10.1109/AICCSA.2018.8612829)
- [3] M. Bezoui, A. Elmoutaouakkil, and A. Beni-Hssane, "Feature extraction of some Quranic recitation using mel-frequency cepstral coefficients (mfcc)," in *International Conference on Multimedia Computing and Systems*, Marrakech, Morocco, Oct. 2016, pp. 127-131. doi: [10.1109/ICMCS.2016.7905619](https://doi.org/10.1109/ICMCS.2016.7905619)
- [4] N. R. R. Rashid, I. Venkat, F. Damanhoori, N. Mustafa, W. Husain, and A. T. Khader, "Towards automating the evaluation of holy Quran recitations: a pattern recognition perspective," in *2013 Taibah University International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, Madinah, Saudi Arabia, Dec. 2013, pp. 424-428. doi: [10.1109/NOORIC.2013.88](https://doi.org/10.1109/NOORIC.2013.88)
- [5] Y. Afrillia, H. Mawengkang, M. Ramli, F. Fadlisyah, and R. P. Fhonna, "Performance measurement of mel frequency cepstral coefficient (mfcc) method in learning system of al-quran based in naghama pattern recognition," *Journal of Physics: Conference Series*, vol. 930, no. 1, pp. 1-6, 2017. doi: [10.1088/1742-6596/930/1/012036](https://doi.org/10.1088/1742-6596/930/1/012036)
- [6] A. Mustofa, "Sistem pengenalan penutur dengan metode mel-frequency wrapping," *Jurnal Teknik Elektro*, vol. 7, no. 2, pp. 88-96, 2007. doi: [10.9744/jte.7.2.88-96](https://doi.org/10.9744/jte.7.2.88-96)
- [7] N. W. Arshad et al., "Makhraj recognition for al-quran recitation using mfcc," *IJIIP: International Journal of Intelligent Information Processing*, vol. 4, no. 2, pp. 45-53, 2013.
- [8] T. Nasution, "Metoda mel frequency cepstrum coefficients (mfcc) untuk mengenali ucapan pada bahasa indonesia," *SATIN - Sains dan Teknologi Informasi*, vol. 1, no. 1, pp. 22-31, 2012.
- [9] O. Nurdiana, J. Jumadi, and D. Nursantika, "Perbandingan metode *cosine similarity* dengan metode jaccard similarity pada aplikasi pencarian terjemah al-qur'an dalam bahasa Indonesia," *Jurnal Online Informatika*, vol. 1, no. 1, pp. 59-63, 2016. doi: [10.15575/join.v1i1.12](https://doi.org/10.15575/join.v1i1.12)
- [10] M. R. Maarif and C. Hadi, "Implementasi cosine similarity dalam aplikasi pencarian ayat al-quran berbasis android," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 6, no. 2, pp. 71-79, 2017.
- [11] I. Wijayanto, B. Hidayat, and Suyanto, "Pemisahan suara musik instrumental menggunakan metode independent component analysis dan pemodelan autoregressive moving average," in *Konferensi Nasional Sistem dan Informatika*, Bali, Indonesia, Nov. 2011, pp. 61-68.
- [12] R. Wulandari, A. Izzuddin, and T. Asrori, "Pengenalan ucapan menggunakan algoritma back propagation," *Energy*, vol. 6, no. 1, pp. 28-36, 2016.
- [13] K. Liu et al., "Endpoint detection of distributed fiber sensing systems based on STFT algorithm," *Optics & Laser Technology*, vol. 114, pp. 122-126, 2019. doi: [10.1016/j.optlastec.2019.01.036](https://doi.org/10.1016/j.optlastec.2019.01.036)
- [14] N. N. Lokhande, N. S. Nehe, and P. S. Vikhe, "Voice activity detection algorithm for speech recognition applications," *International Journal of Computer Applications*, vol. ICCIA, no. 6, pp. 5-7, 2012.
- [15] I. McLoughlin, *Applied science and audio processing with matlab examples*. Cambridge University Press, 2009.
- [16] H. Erokyar, "Age and gender recognition for speech applications based on support vector machines," master thesis, University of South Florida, 2014.
- [17] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann Publishers, 2016.