

DETEKSI KESALAHAN HURUF HIJAYYAH MENGGUNAKAN METODE MEL CEPSTRUM FREQUENCY CEPSTRUM COEFFICIENT

FAIL DETECTION OF HIJAYYAH LETTER PRONOUNCIATION USING MEL FREQUENCY CEPSTRUM COEFFICIENT METHOD

Muh Khaerul Amri S.P¹, Drs. Suprayogi, M.T.², Hertiana Bethaningtyas Dyah K., S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Fisika, Fakultas Teknik Elektro, Universitas Telkom

khaerulamri98@gmail.com¹, hertiana@telkomuniversity.ac.id², spiyogi@yahoo.co.id³

Abstrak

Speech Recognition (SR) adalah teknik untuk mengubah dari informasi *speech* ke *text*. *Mel Frequency Cepstrum Coefficient* sebagai salah satu teknik *Speech Recognition* memiliki prinsip seperti pendengaran manusia. Metode MFCC penting sebab makhraj huruf hijayyah dalam proses pelafalan menggunakan organ *articulator* dan *vocal tract* yang mampu dikenali dengan baik oleh pendengaran manusia. Penelitian terkait diperoleh 12 huruf hijayyah yang memiliki tingkat kesulitan tinggi dalam pelafalan yaitu /ظ/, /ط/, /ض/, /ص/, /ش/, /ذ/, /خ/, /ح/, /ت/, /ق/, /غ/, /ع/. Perkembangan sarana belajar terbaru hanya sebatas memandu orang untuk membaca dengan baik, bukan mengoreksi bacaan sehingga dibutuhkan media layak untuk membantu proses belajar huruf hijayyah. Untuk keperluan akurasi sistem maka diuji penggunaan 3, 6, 9, atau 12 *channel (filter)* MFCC. Penggunaan *channel* dengan akurasi pengujian data uji tertinggi digunakan sebagai acuan sistem.

Kata Kunci: *Speech Recognition, Mel Frequency Cepstrum Coefficient, /ظ/, /ط/, /ض/, /ص/, /ش/, /ذ/, /خ/, /ح/, /ت/, /ق/, /غ/, /ع/, channel (filter)*

Abstract

Speech Recognition (SR) is a technique for converting from speech information to text. *Mel Frequency Cepstrum Coefficient* as one of the techniques of *Speech Recognition* has principles such as human hearing. MFCC method is important because makhraj letters hijayyah in the process of pronunciation using *articulator* and *vocal tract* that can be recognized by human hearing. Related research obtained 12 letters hijayyah which have high difficulty level in pronunciation that is, /ظ/, /ط/, /ض/, /ص/, /ش/, /ذ/, /خ/, /ح/, /ت/, /ق/, /غ/, /ع/. The development of the latest learning facilities is limited to guide people to read well, not to correct the reading so that needed media worth to help the learning process hijayyah letters. For the purpose of system accuracy it is tested using 3, 6, 9, or 12 *channel (filter)* MFCC. The use of the *channel* with the highest test data testing accuracy is used as a system reference.

Key Words : *Speech Recognition, Mel Frequency Cepstrum Coefficient, /ظ/, /ط/, /ض/, /ص/, /ش/, /ذ/, /خ/, /ح/, /ت/, /ق/, /غ/, /ع/, channel (filter)*

1. Pendahuluan

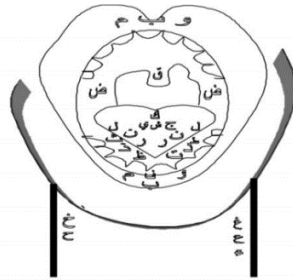
Umat Muslim mampu membaca Al-Quran, namun tidak semua umat Muslim mampu melafalkan Al-Quran sesuai kaidah yang benar berdasarkan ilmu *tajwid* dan *makhraj* [1]. Organ *articulator* dan *vocal tract* (*source & filter*) berperan dalam pelafalan huruf hijayyah [2]. Sehingga SR mampu menangkap informasi pada lafal yang dihasilkan organ *articulator* dan *vocal tract* yang nantinya digambarkan *vector acoustic* [3].

Mel Frequency Cepstrum Coefficient (MFCC) adalah metode SR yang memiliki prinsip kerja telinga manusia. Prinsip ini penting dalam membedakan setiap lafal huruf hijayyah berdasarkan organ *articulator* dan *vocal tract* [4]. Perkembangan media belajar huruf hijayyah hanya terbatas sampai pemandu baca huruf, sehingga dibutuhkan media yang mampu mengoreksi penyebutan lafal huruf hijayyah.

2. Dasar Teori

2.1 Pembentukan Suara

Source-filter speech theory menggambarkan proses pembentukan suara manusia. *Larhynx* (laring) berperan sebagai *source* dalam membentuk suara dan *parhynx* berperan sebagai filter dalam meresonansi aliran udara. *Source-filter* dapat membentuk suara dengan bantuan organ-organ *articulator*. *Larhynx*, *parhynx*, *nasal cavity*, dan *oral cavity* disebut sebagai *vocal tract* pada proses pembentukan huruf hijayyah.



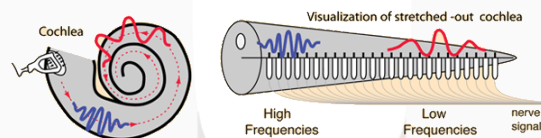
Gambar 2.1 Makhraj Hijayyah Pada Vocal Tract [5]

2.2 Makhraj Sebagai Tempat Keluar Huruf Hijayyah

Hak keluar huruf tersebut dibedakan berdasarkan organ artikulator yang dipakai dalam menyebutkan huruf. Organ-organ yang digunakan dalam makhraj huruf yaitu mulut, lidah, bibir, dan gigi yang termasuk dalam *nasal cavity*, tenggorokan dan pangkal hidung yang termasuk *parhynx (filter)*. Penelitian terkait penyebutan huruf hijayyah menemukan 12 huruf hijayyah memiliki perbedaan dengan huruf pembanding Indonesia sehingga membuat orang Indonesia sulit dalam proses penyebutan huruf tersebut [6]. Berikut 12 huruf tersebut, keterangan makhraj, dan kesalahan umum yang sering terjadi dalam penyebutan pada tabel 2.1.

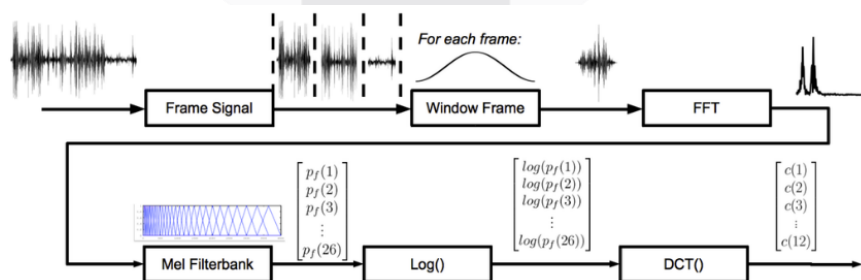
2.3 Mel Frequency Cepstrum Coefficient (MFCC)

Mel Frequency Cepstrum Coefficient adalah salah satu metode dalam *speech recognition* secara fungsi mengikuti karakteristik persepsi pendengaran manusia yang tidak linear namun logaritmik. Didasari pada salah satu organ dalam telinga yaitu koklea yang bergetar pada tempat berbeda pada bagian koklea



Gambar 2.2 Koklea [7]

Hasil dari proses adalah *Cepstral Coefficients*, *Cepstrum* adalah kebalikan dari *Spectrum*. Pada tugas akhir berikut, digunakan 12 *channel/filter* masing-masing berisi 12 kombinasi besaran energi *cepstral* yang membedakan satu huruf dengan huruf lain.



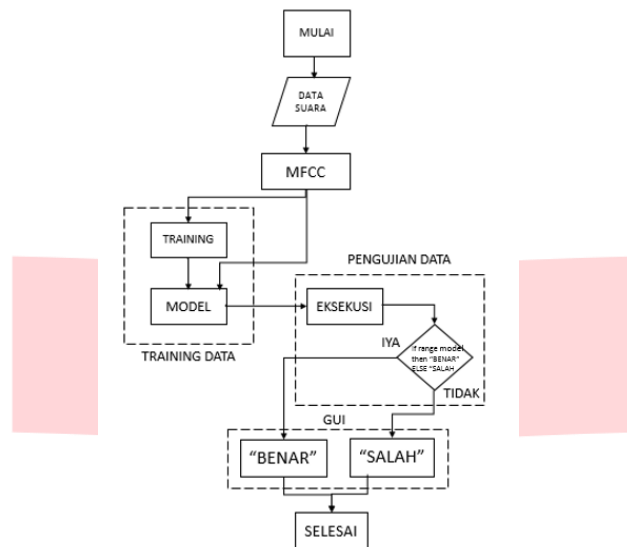
Gambar 2.3 Skema MFCC [8]

Proses framing sinyal suara yang memiliki durasi 0,5 – 1,0 detik dipotong menjadi deretan detik-detik yang lebih singkat 25- 50 ms. Windowing dilakukan dengan mengalikan setiap frame dengan fungsi window hamming dalam mengantisipasi distorsi pada ujung-ujung sinyal hasil framing. FFT (Fast Fourier Transform) dilakukan pada setiap frame untuk mengambil spectrum frekuensi yang akan filter pada Mel filter bank, kemudian proses

logarisasi (log) mengambil spectral envelope (bagian spectrum yang menjadi puncak). Terakhir dilakukan proses invers fourier pada DCT (Discrete Cosinus Transform) untuk mengubah menjadi domain waktu.

3. Pembahasan

3.1 Pengumpulan Data Latih & Uji



Gambar 3.1 Diagram Alur Pendeteksi Huruf Hijayyah

Data latih (*training*) dalam bentuk suara dan tersimpan dalam format .wav sebanyak 900 data suara yang digunakan dalam proses pelatihan data untuk memperoleh model per masing-masing huruf. Data uji dalam bentuk suara dan tersimpan dalam format .wav sebanyak 600 data suara dan digunakan untuk menguji model yang diperoleh dari pelatihan data. Durasi per huruf sepanjang 1 detik.

Pada diagram alir di atas terlihat bagaimana proses memperoleh nilai MFCC pada proses ekstraksi ciri, untuk 900 data yang digunakan sebagai model acuan huruf hijayyah dengan standarisasi partisipan khusus. 600 data uji dengan standarisasi partisipan umum akan dikenali menggunakan metode if then else logic, hasil akhir akan ditampilkan dalam unit interface pada matlab.

No	CH1	CH2	CH3	CH4	CH5	CH6	CH7	CH8	CH9	CH10	CH11	CH12
1	0.7141	-0.18994	0.1678	-0.1145	0.1535	-0.1846	-0.0542	0.0354	0.0230	-0.0332	-0.0348	-0.0156
2	0.562484	-0.76899	-0.2049	-0.3643	0.076182	-0.4022	-0.1619	0.18794	0.230842	0.093891	0.008437	0.005844
...	...	...	...	...	...	...	...	...	...	...	...	...
34	0.376216	-0.52626	-0.3651	-0.5003	0.416011	-0.2196	-0.2037	0.194326	-0.0188	0.03839	0.011676	0.003223
35	0.533532	-0.3856	-0.4983	-0.5373	0.389591	-0.2411	-0.0948	0.167602	0.134134	0.042161	0.033272	0.003882
Deviasi	0.298122	0.288288	0.362556	0.226931	0.270101	0.16234	0.126508	0.109311	0.09227	0.062719	0.039237	0.009146

Tabel 3.2 35 Data MFCC Huruf Ayn

Pada tabel 3.2 Terdapat nilai deviasi pada masing-masing *channel* (CH.1 – 12). Nilai standar deviasi tersebut merepresentasikan seberapa besar simpangan data yang ada pada masing-masing sinyal, semakin besar deviasi maka semakin besar simpangan data yang ada data kumpulan data uji yang dijadikan sebagai model. Penelitian yang dilakukan pada tugas akhir berikut adalah menentukan kombinasi *channel* urutan berapa yang menghasilkan akurasi yang terbaik.

Pada gambar 3. Diagram alur pendeteksi huruf hijayyah, pengambilan keputusan dilakukan menggunakan metode *if then else logic* pada *software* Matlab. Nilai *cepstrum/channel* dari data uji yang berada di luar rentang nilai *cepstrum/channel* model maka akan dikategorikan sebagai data penyebutan salah pada GUI (*Graphic Unit Interface*).

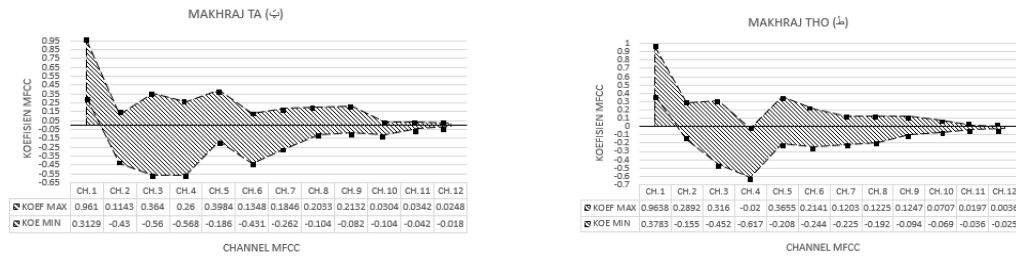
3.2 Hasil Training Data

URUTAN DEVIASI	1	2	3	4	5	6	7	8	9	10	11	12
CHANNEL	CH.12	CH.11	CH.10	CH.9	CH.8	CH.7	CH.6	CH.5	CH.4	CH.3	CH.1	CH.2
NILAI DEVIASI	0.00599	0.02336	0.04325	0.06168	0.08058	0.10379	0.13029	0.14068	0.16146	0.20126	0.20361	0.21707
RANGE MFCC	-0.01573	-0.0714	-0.1077	-0.05869	-0.08705	-0.28625	-0.29006	-0.18303	-0.43016	-0.56028	0.11428	-0.88382
	0.01286	0.04716	0.07892	0.1678	0.19665	0.23883	0.23884	0.56301	0.17861	0.25014	0.8974	0.28019

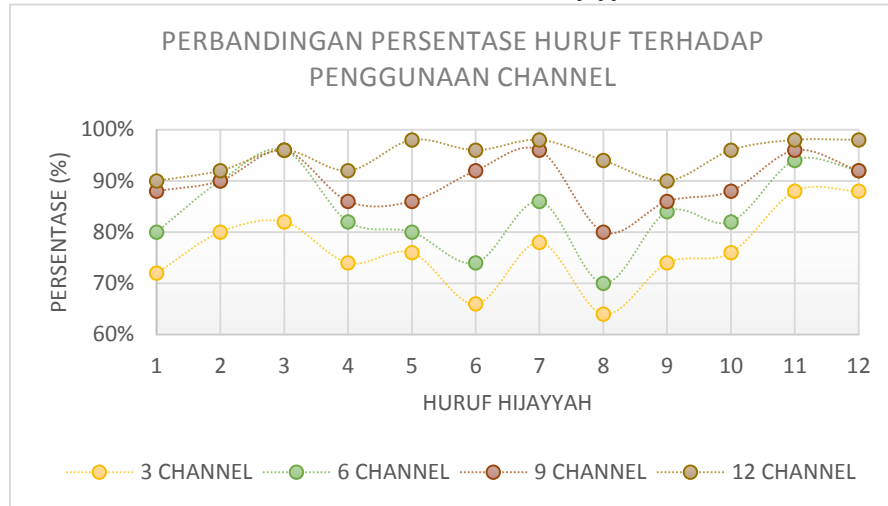
Tabel 3.3 Model Huruf Ayn

Pada tabel 4.3 pada pengujian tingkat ketepatan data uji untuk penggunaan 3 *channel* bernilai deviasi terkecil adalah CH.12 bernilai 0.00599, CH.11 bernilai 0.02336 & CH.10 bernilai 0.04325, kemudian pada penggunaan 6 *channel* adalah CH.12 bernilai 0.00599, CH.11 bernilai 0.02336, CH.10 bernilai 0.04325, CH.9 bernilai 0.06168, CH.8 bernilai 0.08058, & CH.7 bernilai 0.13079, kemudian pada penggunaan 9 *channel* CH.12 bernilai 0.00599, CH.11 bernilai 0.02336, CH.10 bernilai 0.04325, CH.9 bernilai 0.06168, CH.8 bernilai 0.08058, CH.7 bernilai 0.13079, CH.6 bernilai 0.13029, CH.5 bernilai 0.14068, & CH.4 bernilai 0.16146.





Gambar 3.4 Model MFCC 12 Huruf Hijayyah 12 Channel

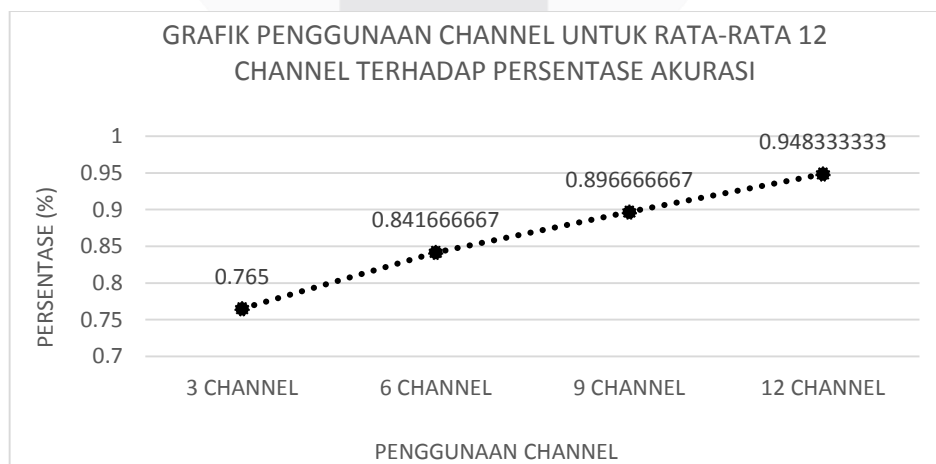


Gambar 3.5 Grafik Perbandingan Persentase 12, 9, 6, dan 3 Channel MFCC

Terlihat grafik sebaran persentase untuk masing-masing huruf hijayyah sejumlah 12 huruf, dengan penggunaan jumlah channel berbeda berdasarkan nilai deviasi dari data latih huruf. Pada sumbu horizontal grafik, nomor 1 = Ayn, 2 = Dlod, 3 = Dzal, 4 = Dzo, 5 = Ghin, 6 = Ha, 7 = Kha, 8 = Qof, 9 = Shod, 10 = Syin, 11 = Ta, & 12 = Tho. Penggunaan *n channel* digambarkan dalam titik-titik berwarna berbeda. Tabel dia atas memperlihatkan hasil pengujian data untuk penggunaan 3,6,9, dan 12 channel huruf hijayyah.

Persentase terkecil pada penggunaan channel ke-3 bernilai deviasi terkecil Terdapat pada huruf qof sebesar 64%. Persentase terbesar terdapat pada huruf ta dan tho sebesar 88%. Dalam penggunaan channel ke-6 bernilai deviasi terkecil, persentase terkecil terdapat pada huruf ha sebesar 74%. Persentase terbesar terdapat pada huruf ta sebesar 94%.

Dalam penggunaan channel ke-9 bernilai deviasi terkecil, persentase terkecil terdapat pada huruf qof sebesar 80%. Untuk persentase terbesar diperoleh huruf dzal, kha, dan ta sebesar 96%. Dalam penggunaan 12 channel lengkap, persentase terkecil terdapat pada huruf ayn dan shod sebesar 90%. Persentase terbesar terdapat pada huruf ghin, ta, dan tho sebesar 98%.



Grafik 3.6 Persentase Rata-Rata Akurasi 12 Huruf Hijayyah

Rata-rata persentase tingkat ketepatan menguji data *testing* (uji) yang diperoleh dari penggunaan 3 *channel* sebesar 76.5%, 6 *channel* sebesar 84.16%, 9 *channel* sebesar 89.66%, dan 12 *channel* sebesar 94.83%. Sehingga banyak penggunaan *channel* yang digunakan dalam perangkat lunak pendeteksi huruf hijayyah adalah penggunaan 12 *channel* dengan pertimbangan hasil *testing* data uji.

Kesimpulan

Dari hasil penelitian yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut.

1. Penerapan algoritma MFCC dari tahap framing hingga DCT (Discrete Continues Transform) pada data latih dan data uji menghasilkan besaran vector acoustic berupa 12 deret nilai. Diperoleh range nilai MFCC/channel dan standar deviasi dari pelatihan data untuk proses pencocokan data. Sehingga diperoleh persentase akurasi terbaik oleh huruf Ta dan Tho sebesar 98% untuk penggunaan 12 *channel*. Persentase akurasi terendah oleh huruf Qof sebesar 64% menggunakan 3 *channel*.
2. Penerapan teknik pencocokan dan pengambil keputusan data *if the else logic*. Teknik tersebut digunakan dengan pertimbangan model data yang diperoleh diambil dari range nilai MFCC/channel yang memiliki ambang batas nilai cepstral/channel. Hasil akurasi rata-rata pengujian 3 *channel* sebesar 76.5%, 6 *channel* sebesar 84.16%, 9 *channel* sebesar 89.66%, dan 12 *channel* sebesar 94.83%
3. Berdasarkan persentase akurasi dari penggunaan 3 *channel*, 6 *channel*, 9 *channel*, dan 12 *channel* yang dipilih berdasarkan besar urutan nilai standar deviasi *channel*, maka persentase tertinggi rata-rata 12 huruf hijayyah diperoleh oleh penggunaan 12 *channel* MFCC sebesar 94,83%.

Daftar Pustaka

- [1]. Arshad, N. W., Sukri, S. M., Muhammad, L. N., Ahmad, H., Hamid, R., Naim, F., & Naharuddin, N. A. (2013, June). Journal. *Makhkraj Recognition for Al-Quran Recitation using MFCC*, 45-52.
- [2]. Sefiana, E. (2011). Kompetensi Fonologis Anak. In E. Sefiana, *Kompetensi Fonologis Anak* (p. 14). Bandung: Universitas Pendidikan Bandung.
- [3]. Rahman, Y. H., Adiwijaya, & Maharani, W. (2012). Journal. *Pengenalan Huruf Hijayyah Berbasis Speech Recognition Menggunakan Hidden Markov Model (HMM)*, 1.
- [4]. Fauzi, R. M. (2013). *Pengenalan Ucapan Huruf Hijaiyah Menggunakan Mel Frequency Cepstral Coefficients (MFCC) dan Hidden Markov Model (HMM)*. Bandung: Institute Tekonologi Telkom.
- [5]. Subali, M., Djasri, D., & Alawiyah, N. (2014). Frekuensi Forman Sebagai Model Akustik Tabung Sederhana Dari Vocal Tract. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, A-41.
- [6]. Hanapi, A. (2015). Buku Panduan Belajar Peserta LKP TAR-Q. In A. Hanapi, *Materi Praktis Tahsin Tilawah Jilid 2* (pp. 5-11). Bandung: Tar-Q Press.
- [7]. *Sensitive Of Human Ear*. (n.d.). Retrieved from <http://hyperphysics.phy-astr.gsu.edu/http://hyperphysics.phy-astr.gsu.edu/hbase/Sound/earsens.html>
- [8]. K.K, B., & Kopparapu, S. (2014). Modified Mel Filter Bank to Compute MFCC of Subsampled Speech. *The Computing Reserach Repository*, 1-4.