

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Penguasaan bahasa asing memiliki peran penting dalam kehidupan modern, baik dalam konteks komunikasi lintas budaya, pengembangan ilmu pengetahuan, maupun kebutuhan keagamaan. Bahasa Arab merupakan salah satu bahasa asing yang memiliki kedudukan strategis secara global, dengan jumlah penutur yang besar serta perannya dalam berbagai aspek kehidupan, seperti budaya, pendidikan, dan hubungan internasional [1]. Di Indonesia, bahasa Arab dipelajari secara luas, terutama dalam lingkungan pendidikan berbasis Islam, mulai dari tingkat dasar hingga perguruan tinggi. Namun, berbeda dengan bahasa asing pada umumnya, pembelajaran bahasa Arab dalam konteks keagamaan tidak hanya menekankan pada pemahaman makna, tetapi juga ketepatan pelafalan pada tingkat fonetik yang rinci [2]. Ketepatan pelafalan ini diatur dalam ilmu tajwid, khususnya pada aspek makhraj yang membahas tempat keluarnya bunyi huruf hijaiyah dari organ ucap.

Makhraj merupakan aspek fundamental dalam ilmu *tajwid* yang menentukan ketepatan tempat keluarnya huruf *hijaiyah* saat melafalkan ayat Al-Qur'an. Ketepatan *makhraj* berpengaruh langsung terhadap kejelasan bunyi dan makna bacaan, sehingga kesalahan artikulasi dapat mengubah karakter fonetik suatu huruf [3]. Namun, berbagai studi menunjukkan bahwa kemampuan pelafalan *makhraj*, khususnya pada pembelajar *non-native* bahasa Arab, masih berada pada tingkat yang belum optimal. Penelitian *Student speaking skill assessment: Techniques and results* menunjukkan bahwa keterampilan berbicara bahasa Arab siswa berada pada kategori cukup dengan nilai rata-rata 3,09 pada skala Likert 1–5, di mana nilai tersebut berada di atas batas minimal tetapi belum mencapai kategori baik [4]. Skala Likert ini menginterpretasikan nilai 1 sebagai sangat rendah, 3 sebagai cukup, dan 5 sebagai sangat baik [5]. Temuan tersebut mengindikasikan bahwa meskipun pelafalan dasar telah tercapai, masih terdapat kelemahan pada aspek ketepatan artikulasi, termasuk penguasaan *makhraj* dan kejelasan vokal, yang memerlukan metode evaluasi yang lebih objektif dan terukur.

Salah satu tantangan utama dalam mempelajari makhraj huruf hijaiyah adalah adanya pasangan huruf yang secara fonetik sangat mirip namun berbeda titik artikulasinya, sehingga mudah tertukar oleh penutur non-Arab. Sebagai

contoh, huruf *ha* dan *ha* keduanya menghasilkan bunyi yang bagi penutur bahasa Indonesia terdengar seperti bunyi /h/, namun berbeda secara fonetik: *ha* dihasilkan dari tenggorokan bagian tengah (*wasath al-halq*) dengan gesekan udara yang lebih kuat, sedangkan *ha* dihasilkan dari tenggorokan bagian bawah (*aqsha al-halq*) dengan bunyi yang lebih ringan [6]. Contoh lain adalah huruf *sin*, *shad*, dan *tsa* yang ketiganya menghasilkan bunyi desis mirip /s/ namun masing-masing memiliki posisi lidah dan bentuk rongga mulut yang berbeda. Kesalahan melafalkan pasangan-pasangan huruf seperti ini dapat mengubah makna kata dalam bahasa Arab, sehingga ketepatan makhraj bukan sekadar persoalan estetika bacaan melainkan berpengaruh pada substansi pemahaman.

Secara tradisional, pembelajaran dan evaluasi pelafalan makhraj dilakukan melalui metode tatap muka atau *talaqqi* yang sangat bergantung pada bimbingan langsung guru. Meskipun metode ini efektif dalam menjaga ketepatan bacaan, keterbatasan waktu, akses terhadap pengajar yang kompeten, serta sifat penilaian yang subjektif menjadi tantangan tersendiri, terutama dalam konteks pembelajaran mandiri [7]. Penilaian lisan yang bergantung pada persepsi auditoris guru berpotensi menimbulkan variasi interpretasi terhadap tingkat ketepatan artikulasi, khususnya pada bunyi-bunyi yang memiliki kemiripan fonetik seperti yang telah dicontohkan di atas. Oleh karena itu, diperlukan suatu sistem evaluasi yang mampu memberikan pengukuran akurasi secara lebih objektif dan terstandar. Perkembangan teknologi kecerdasan buatan memungkinkan pengembangan sistem evaluasi berbasis pengolahan sinyal suara yang tidak hanya mengenali bunyi, tetapi juga mengukur tingkat kesesuaian pelafalan terhadap referensi yang benar. Dalam konteks ini, sistem berbasis *speech processing* dapat berperan sebagai *decision support system* yang membantu proses pembelajaran dengan menyediakan umpan balik kuantitatif terhadap akurasi artikulasi makhraj, khususnya pada pelafalan huruf hijaiyah berharakat fathah.

Seiring dengan berkembangnya teknologi kecerdasan buatan, pendekatan berbasis *Machine Learning* dan *Deep Learning* mulai banyak diterapkan untuk membantu evaluasi pelafalan secara otomatis dan objektif. Salah satu teknik ekstraksi ciri audio yang paling umum digunakan adalah *Mel-Frequency Cepstral Coefficients* (MFCC) karena kemampuannya merepresentasikan karakteristik sinyal suara yang mendekati persepsi pendengaran manusia. Penelitian [8] menerapkan MFCC yang dikombinasikan dengan *Convolutional Neural Network* (CNN) untuk mendeteksi *makhraj* huruf-huruf halqi dalam pelafalan Al-Qur'an dan berhasil mencapai akurasi sebesar 89,39%, menunjukkan efektivitas CNN dalam

mengklasifikasikan ciri fonetik berbasis MFCC. Dalam lingkup yang lebih dekat, penelitian dari institusi yang sama menunjukkan bahwa CNN dengan arsitektur EfficientNet-B7 berhasil diterapkan untuk pengenalan kosakata bahasa Arab dari citra objek dengan akurasi 90% serta *precision*, *recall*, dan *F1-score* rata-rata 94,6% [9], membuktikan efektivitas pendekatan *deep learning* pada domain bahasa Arab dan mendukung penerapannya pada evaluasi pelafalan huruf hijaiyah dalam penelitian ini. Di sisi lain, penelitian [10] memanfaatkan kombinasi MFCC dan *Dynamic Time Warping* (DTW) untuk membandingkan pola pelafalan suara, dengan hasil akurasi rata-rata 88,89%, yang menegaskan keandalan DTW dalam mengukur kesesuaian temporal antar sinyal audio. Pendekatan serupa juga diterapkan pada domain yang berbeda, seperti klasifikasi lingkungan suara pada penelitian [11], di mana MFCC digunakan sebagai fitur utama dan CNN dikembangkan sebagai model klasifikasi, menghasilkan peningkatan akurasi dibandingkan model *baseline*. Selain itu, penelitian [12] menunjukkan bahwa pengembangan arsitektur CNN yang lebih optimal mampu meningkatkan performa pengenalan suara hingga 10–25% dibandingkan pendekatan MFCC-CNN konvensional maupun DTW tunggal. Pada domain musik, penelitian berjudul “Penerapan *Convolutional Neural Network* untuk Melakukan Estimasi Pitch pada Rekaman Suara Penyanyi” menunjukkan bahwa kualitas pembawaan karya musik sangat ditentukan oleh ketepatan intonasi, yang merepresentasikan akurasi *pitch* dari nada yang dihasilkan. Dalam penelitian tersebut, CNN satu dimensi digunakan untuk melakukan estimasi *pitch* dari rekaman suara nyanyian, kemudian hasil estimasi dibandingkan dengan template menggunakan DTW untuk mengukur tingkat kemiripan. Melalui kombinasi CNN dan DTW, sistem mampu menghasilkan rata-rata akurasi tertinggi sebesar 97,425% [13].

Keberhasilan pendekatan CNN dalam mempelajari representasi fitur suara serta kemampuan DTW dalam mengukur kesesuaian pola temporal menunjukkan potensi besar integrasi kedua metode tersebut dalam sistem evaluasi berbasis sinyal audio. Jika pada penelitian tersebut CNN digunakan untuk estimasi *pitch* dan DTW untuk mengukur ketepatan intonasi, maka pendekatan serupa secara konseptual dapat diadaptasi untuk mengevaluasi akurasi artikulasi makhraj huruf hijaiyah, khususnya pada pelafalan berharakat fathah. Perbedaannya terletak pada fokus evaluasi: bukan pada kesesuaian tinggi-rendah nada (*pitch*), melainkan pada ketepatan karakteristik spektral dan pola artikulatoris yang membedakan satu makhraj dengan makhraj lainnya.

Dalam konteks pengolahan sinyal suara, MFCC telah menjadi

teknik ekstraksi ciri yang sangat umum digunakan karena kemampuannya merepresentasikan karakteristik spektral suara yang mendekati persepsi pendengaran manusia. Berbagai penelitian menunjukkan bahwa kombinasi MFCC dan CNN efektif dalam klasifikasi bunyi fonetik, sementara DTW andal dalam mengukur kesesuaian pola waktu antar sinyal suara [8, 10, 11, 12, 13]. Namun, kajian yang secara spesifik mengintegrasikan MFCC, CNN, dan DTW sekaligus dalam satu sistem terpadu untuk mengevaluasi akurasi artikulasi makhraj huruf hijaiyah—khususnya pada pelafalan vokal fathah—masih sangat terbatas. Hal ini menjadi peluang penelitian untuk mengembangkan dan membandingkan dua pendekatan evaluasi yang berbeda: pendekatan berbasis klasifikasi spektral menggunakan MFCC dan CNN, serta pendekatan berbasis pencocokan temporal menggunakan MFCC dan DTW. Dengan membandingkan kedua pendekatan secara empiris pada domain yang sama, penelitian ini diharapkan dapat memberikan gambaran objektif tentang kelebihan dan keterbatasan masing-masing metode dalam mengevaluasi akurasi artikulasi makhraj huruf hijaiyah berharakat fathah, sekaligus mendukung pengembangan sistem evaluasi pelafalan yang lebih presisi dan adaptif.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan sistem evaluasi akurasi artikulasi makhraj huruf hijaiyah berharakat fathah berbasis ekstraksi fitur MFCC, model *Convolutional Neural Network* (CNN), dan *Dynamic Time Warping* (DTW)?
2. Bagaimana *distance score* yang dihasilkan Metode B (MFCC-DTW) dapat mendukung evaluasi ketepatan artikulasi makhraj huruf hijaiyah berharakat fathah yang dihasilkan oleh *confusion matrix* Metode A (MFCC-CNN)?

1.3 Batasan Permasalahan

Agar menjaga ketetapan arah penelitian, diterapkan beberapa batasan permasalahan sebagai berikut:

1. Data penelitian berupa rekaman suara huruf hijaiyah dengan harakat fathah dalam format .wav dengan durasi kurang dari 1 detik per sampel.
2. Penelitian difokuskan pada evaluasi akurasi artikulasi makhras huruf hijaiyah berharakat fathah, bukan pada deteksi hukum tajwid lainnya seperti hukum nun mati atau mad.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan sistem evaluasi akurasi artikulasi makhras huruf hijaiyah berharakat fathah menggunakan pendekatan MFCC-CNN dan MFCC-DTW sebagai dua metode evaluasi yang berjalan secara independen.
2. Menganalisis peran *distance score* dari Metode B (MFCC-DTW) dalam mendukung evaluasi ketepatan artikulasi makhras yang dihasilkan *confusion matrix* Metode A (MFCC-CNN), serta mengidentifikasi karakteristik dan keterbatasan masing-masing metode.

1.5 Manfaat Penelitian

Pada bagian ini diuraikan manfaat yang akan diperoleh dari penelitian ini. Penyajiannya dalam bentuk poin-poin, dan manfaat yang disampaikan haruslah sesuai dengan lingkup kerja penelitian.

1. Memberikan kontribusi dalam pengembangan sistem evaluasi pelafalan makhras berbasis kecerdasan buatan yang lebih objektif dan terukur.
2. Mendukung pembelajaran mandiri bahasa Arab dan Al-Qur'an melalui sistem umpan balik otomatis terhadap ketepatan pelafalan huruf hijaiyah berharakat fathah.
3. Menjadi referensi bagi penelitian selanjutnya dalam pengembangan metode integrasi MFCC, CNN, dan DTW pada pengolahan sinyal suara.

1.6 Sistematika Penulisan

Bagian ini berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran. Secara umum, sistematika penulisan laporan adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**
Berisi latar belakang, rumusan masalah, batasan permasalahan, tujuan, manfaat, dan sistematika penulisan.
- **Bab 2 LANDASAN TEORI**
Berisi kajian pustaka dan landasan teori yang mendukung penelitian, termasuk teori makhray, tajwid, pengolahan sinyal suara, MFCC, CNN, dan DTW.
- **Bab 3 METODOLOGI PENELITIAN**
Berisi metodologi penelitian yang menjelaskan alur penelitian, dataset, konfigurasi sistem, preprocessing audio, ekstraksi fitur MFCC, arsitektur CNN, mekanisme DTW, serta prosedur evaluasi.
- **Bab 4 HASIL DAN DISKUSI**
Berisi hasil implementasi dan evaluasi sistem, termasuk hasil pelatihan CNN, hasil evaluasi DTW, analisis perbandingan kedua metode, serta pembahasan temuan utama.
- **Bab 5 KESIMPULAN DAN SARAN**
Berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A