ABSTRAK

Evaluasi bacaan Al-Qur'an pada pembelajaran DIROSA tingkat awal masih bertumpu pada penyimakan langsung oleh ustadz, sementara peserta memerlukan indikator awal ketika berlatih secara mandiri. Penelitian penilaian bacaan Al-Qur'an berbantuan komputer umumnya berfokus pada pengenalan atau deteksi kesalahan yang memerlukan transkripsi, label fonem, segmentasi, atau aturan tajwid tertentu, sehingga bukti empiris mengenai hubungan skor kemiripan berbasis representasi audio WavLM dan *Dynamic Time Warping* (DTW) dengan penilaian ustadz masih terbatas. Penelitian ini bertujuan merumuskan dan mengevaluasi metode komputasional untuk mengukur kemiripan bacaan peserta terhadap bacaan referensi ustadz pada unit frasa pendek tanpa memerlukan transkripsi maupun anotasi fonem.

Penelitian menggunakan pendekatan perancangan metode komputasional dengan evaluasi kuantitatif korelasional. Data terdiri atas 20 frasa yang masing-masing berisi 3 sampai 8 kata, direkam oleh satu ustadz pengajar DIROSA dan dibaca oleh lima peserta tingkat awal, sehingga menghasilkan 100 pasangan audio. Setiap rekaman peserta diberi skor oleh ustadz pada skala 1 sampai 5. Audio diproses melalui konversi menjadi mono, pencuplikan ulang menjadi 16 kHz, pemangkasan hening pada awal dan akhir berbasis *Voice Activity Detection*, penyempurnaan batas sinyal berbasis energi, dan normalisasi amplitudo. Representasi audio diekstraksi dari *layer* 1 sampai *layer* 12 model microsoft/wavlm-base-plus dengan parameter yang dibekukan. Pada setiap *layer*, matriks biaya dibentuk menggunakan *cosine distance* dan kedua urutan representasi disejajarkan menggunakan DTW dengan batas pita Sakoe-Chiba sebesar 0,1. Jarak hasil penyelarasan dinormalisasi berdasarkan panjang *warping path*, kemudian dipetakan menjadi skor kemiripan 0 sampai 100 melalui fungsi *sigmoid*. Metode dievaluasi melalui verifikasi fungsional, korelasi Spearman sebagai analisis utama, dan korelasi Pearson sebagai analisis pendukung terhadap skor ustadz.

Hasil verifikasi menunjukkan bahwa seluruh tahapan metode berjalan sesuai dengan rancangan. Skor kemiripan pada seluruh *layer* berkorelasi positif dan signifikan secara statistik dengan skor ustadz. Korelasi tertinggi diperoleh pada *layer* 10 dengan Spearman rho sebesar 0,7635 dan Pearson r sebesar 0,7460, yang termasuk kategori kuat. *Layer* 9, *layer* 11, dan *layer* 12 juga menghasilkan nilai korelasi yang berdekatan, yang mengindikasikan bahwa representasi pada *layer* menengah-akhir hingga akhir lebih selaras dengan pola penilaian ustadz dibandingkan *layer* awal. Rata-rata skor penilaian ustadz per pertemuan bervariasi, dengan nilai terendah pada Pertemuan 19 sebesar 1,20 dan nilai tertinggi pada Pertemuan 4 dan Pertemuan 8 sebesar 2,60. Penelitian ini berkontribusi berupa konfigurasi dan evaluasi awal metode WavLM-DTW sebagai dasar komputasional bagi alat bantu evaluasi formatif, bukan sebagai diagnosis tajwid otomatis atau pengganti penilaian ustadz.

Kata Kunci: WavLM, *Dynamic Time Warping*, DIROSA, kemiripan bacaan Al-Qur'an, representasi audio

ABSTRACT

Evaluation of Qur'anic recitation in early-stage DIROSA learning still relies on direct listening by an ustadz, while learners need an initial indicator when practising independently. Computer-assisted Qur'anic recitation assessment research generally focuses on recognition or error detection methods requiring transcription, phoneme labels, segmentation, or explicit tajwid rules, leaving limited empirical evidence on the relationship between WavLM-DTW-based similarity scores and an ustadz's ratings. This study therefore formulates and evaluates a computational method for measuring the similarity between learners' recitations and an ustadz's reference recitations at the short-phrase level without requiring transcription or phoneme annotation.

The study adopted a computational method design with quantitative correlational evaluation. The dataset comprised 20 phrases, each containing 3 to 8 words, recorded by one DIROSA ustadz and recited by five early-stage learners, yielding 100 audio pairs. Each learner recording was rated by the ustadz on a scale from 1 to 5. Audio pre-processing included mono conversion, resampling to 16 kHz, Voice Activity Detection-based endpoint silence trimming, energy-based boundary refinement, and amplitude normalisation. Audio representations were extracted from layers 1 to 12 of the frozen microsoft/wavlm-base-plus model. At each layer, a cost matrix was constructed using cosine distance, and the two representation sequences were aligned using DTW with a Sakoe-Chiba band ratio of 0.1. The alignment distance was normalised by the warping-path length and mapped to a similarity score ranging from 0 to 100 using a sigmoid function. The method was evaluated through functional verification, Spearman correlation as the primary analysis, and Pearson correlation as a supporting analysis with the ustadz's ratings.

Functional verification confirmed that all stages of the method operated as designed. Similarity scores from all layers correlated positively and significantly with the ustadz's ratings. The highest correlation was obtained at layer 10, with a Spearman rho of 0.7635 and a Pearson r of 0.7460, both indicating a strong relationship. Layers 9, 11, and 12 produced closely comparable correlation values, indicating that representations from the middle-to-late and final layers aligned more closely with the ustadz's assessment pattern than representations from the early layers. Average ustadz ratings varied across meetings, with the lowest value at Meeting 19 (1.20) and the highest values at Meetings 4 and 8 (2.60). This study contributes an initial configuration and evaluation of the WavLM-DTW method as a computational basis for a formative assessment aid, not as an automatic tajwid diagnosis tool or a replacement for an ustadz's assessment.

Keywords: *WavLM, Dynamic Time Warping, DIROSA, Qur'anic recitation similarity, audio representation*