Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 2. November 2023

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Konversi Suara Ke Midi Menggunakan *Short Time Fourier Transform* Sebagai *Virtual Midi Controller* Pada *Digital Audio Workstation*

Yoel Samosir[a1], I Ketut Gede Suhartana[a2], I Gusti Ngurah Anom Cahyadi Putra[a3]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Badung, Bali, Indonesia
[1]yoelsamosir@gmail.com
[2]ikg.suhartana@unud.ac.id
[3]anom.cp@unud.ac.id

Abstrak

Musik menjadi bagian tak terpisahkan dari kehidupan sehari-hari, dan permintaannya terus meningkat berkat kemajuan teknologi. Saat ini, produser musik amatir semakin banyak mengandalkan peralatan digital, seperti pengontrol MIDI, untuk menciptakan musik secara independen. Namun, pengontrol MIDI umumnya memiliki harga yang tinggi, dan tidak semua musisi memiliki kemampuan untuk memainkan alat musik piano atau keyboard.

Dalam penelitian ini, dikembangkan sebuah metode konversi suara ke format MIDI yang menggunakan teknik *Short-Time Fourier Transform* (STFT). Metode ini juga mengenali tingkat akurasi pendeteksian nada dan keandalan informasi yang dihasilkan. Data rekaman audio dari berbagai alat musik dan suara manusia digunakan sebagai input dalam sistem berbasis website. Proses STFT diterapkan pada sinyal audio untuk mengidentifikasi dan mengonversi nada menjadi notasi MIDI.

Hasil analisis menunjukkan bahwa metode STFT mampu menghasilkan tingkat akurasi pendeteksian nada yang cukup memadai, mencapai 24.216621%. Beberapa faktor, seperti kualitas audio, parameter STFT, dan pengaturan threshold, ternyata memiliki pengaruh signifikan terhadap hasil konversi. Penggunaan audio yang berkualitas tinggi, pemilihan parameter STFT yang tepat, dan pengaturan threshold yang optimal dapat meningkatkan akurasi pendeteksian nada secara keseluruhan.

Kata Kunci: *Short-Time Fourier Transform, STFT, Midi, Wav, Audio*

1. Pendahuluan

Perkembangan teknologi dan akses mudah terhadap informasi di era saat ini telah mengubah cara orang melakukan berbagai aktivitas, termasuk dalam industri musik. Salah satu perubahan yang signifikan adalah terciptanya para "Produser Kamar Tidur" yang dapat dengan mudah menciptakan musik secara independen menggunakan teknologi digital yang terjangkau. Mereka menggunakan alat musik berbasis pengontrol MIDI dan teknologi studio virtual untuk menciptakan musik yang dapat dipublikasikan secara global.

Meskipun akses ke teknologi semakin mudah, beberapa peralatan seperti pengontrol MIDI masih memiliki harga yang mahal dan sulit dijangkau oleh semua kalangan. Selain itu, proses pembuatan file MIDI juga memerlukan waktu dan usaha yang cukup panjang. Untuk mengatasi kendala ini, penelitian ini bertujuan untuk mengembangkan sebuah pengontrol MIDI berbasis perangkat lunak yang dapat mengkonversi suara menjadi format MIDI. Hal ini diharapkan dapat mengurangi biaya dan meningkatkan efisiensi bagi para musisi dan produser. Untuk mencapai tujuan tersebut, penelitian menggunakan teknik *Short Time Fourier Transform* (STFT) untuk menganalisis suara dalam domain frekuensi. Dengan STFT, suara dapat dipecah menjadi frame waktu kecil dan diubah menjadi representasi domain frekuensi. Teknik ini memungkinkan identifikasi dan pemisahan komponen frekuensi dalam suara yang nantinya dapat dikonversi menjadi format MIDI.

Sebuah penelitian yang dilakukan oleh Pernando pada tahun 2015 dengan judul "Pengembangan MIDI Controller Berbasis *Microcontroller* Dengan Mekanisme Sentuh" oleh Pratama pada tahun 2014, menghasilkan pengembangan pengontrol MIDI yang berbasis mikrokontroler dengan biaya yang lebih terjangkau. Pengontrol MIDI yang dikembangkan memiliki bentuk yang serupa dengan alat musik piano atau keyboard. Namun, penelitian ini menemukan beberapa kendala terkait penggunaan MIDI Controller, yaitu tidak semua musisi memiliki keterampilan dalam memainkan alat musik piano atau keyboard. Kendala ini mengurangi efisiensi penggunaan MIDI Controller dalam mengirim pesan MIDI secara real-time ke aplikasi DAW (*Digital Audio Workstation*). Akibatnya, pesan MIDI yang terkirim mungkin tidak sesuai dengan tempo atau preferensi yang diinginkan oleh musisi atau produser [1].
Dengan mempertimbangkan permasalahan tersebut, peneliti bermaksud untuk mengembangkan sebuah pengontrol MIDI berbasis perangkat lunak yang dapat mengubah suara menjadi format MIDI untuk kemudian dikirim ke aplikasi DAW. Dengan beberapa tujuan Pertama, untuk mengetahui tingkat akurasi pendeteksian nada menggunakan metode *Short-Time Fourier Transform* (STFT). Kedua, untuk mengetahui faktor-faktor yang mempengaruhi akurasi pendeteksian nada. serta ketiga, untuk mengetahui informasi yang dapat diperoleh dan diubah menjadi format MIDI.
Harapannya, dengan menciptakan pengontrol MIDI berbasis perangkat lunak ini, dapat mengurangi biaya produksi hingga mencapai tingkat minimal atau bahkan tanpa biaya sama sekali. Selain itu, peneliti berharap bahwa pengontrol MIDI ini akan memudahkan musisi dan produser dalam menggunakan berbagai alat musik yang mereka miliki. Sebagai tambahan, musisi atau produser dapat menggunakan suara dari instrumen mereka sendiri yang telah diubah menjadi format MIDI dan mengirimkannya ke aplikasi DAW. Setelah proses konversi, suara tersebut dapat diproses dan diadaptasi menjadi berbagai jenis alat musik yang berbeda.

2. Metode Penelitian

2.1 Pengumpulan Data

Pada tahap ini, data yang akan digunakan sebagai input dalam sistem yang telah dirancang dikumpulkan. Jenis data yang digunakan adalah data kualitatif berupa rekaman suara dari berbagai alat musik dan suara manusia dalam format file .wav. Penggunaan file WAV dipilih karena format ini tidak mengalami kompresi saat di *encode*, sehingga semua elemen audio asli tetap terjaga dalam file tersebut. Pengumpulan data dilakukan melalui dua sumber, yaitu data primer yang direkam dari narasumber saat memainkan alat musik dan menyanyi, serta data sekunder yang diperoleh dari situs yang menyediakan file suara alat musik dengan format .wav yang direkam secara profesional. Total data yang digunakan dalam penelitian sebanyak 11 audio dengan frekuensi sampling 44100 Hz, sesuai standar pemrosesan audio umum yang digunakan. Pada Gambar 1 merupakan data audio yang digunakan dalam penelitian.

Gambar 1. Data Penelitian

2.2 Gambaran Umum Sistem

Sistem pada penelitian ini sebuah platform berbasis website yang dirancang untuk melakukan konversi audio ke dalam bentuk MIDI. Metode yang digunakan untuk konversi ini adalah *Short-Time Fourier Transform* (STFT), yang diaplikasikan pada sinyal audio yang diperoleh dari pengguna melalui file WAV yang diunggah. Selain itu, sistem juga memiliki fitur untuk menghitung skor kemiripan antara file asli dan file yang sudah dikonversi. Pengguna dapat

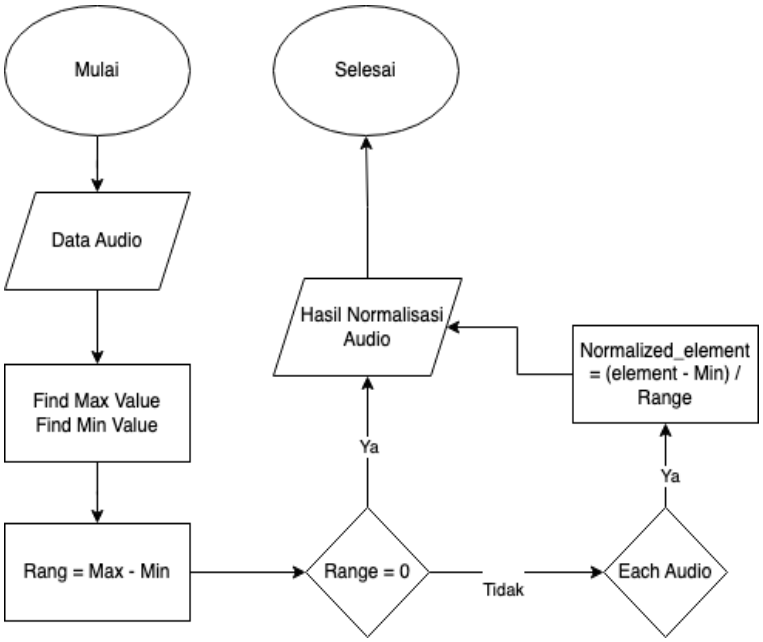
p-ISSN: 2301-5373
e-ISSN: 2654-5101

melakukan pengukuran skor tersebut dengan mengunggah kedua file, yaitu file asli dan hasil konversi.

2.3 Preprocessing

Sebelum melanjutkan analisis lebih lanjut, tahap preprocessing data memiliki peran yang sangat penting dalam proses ini. Salah satu teknik yang digunakan dalam preprocessing data adalah peak normalisasi [2]. Tujuan utama dari peak normalization adalah untuk menghasilkan representasi data yang lebih konsisten dan dapat dibandingkan secara relatif. Dengan melakukan peak normalization, amplitudo puncak dari data audio akan diatur sehingga mencapai level tertentu yang ditentukan sebelumnya. Hal ini membantu menghindari distorsi akibat perbedaan amplitudo yang signifikan antara data, sehingga memungkinkan perbandingan dan analisis yang lebih akurat.

Proses *peak normalization* dimulai dengan mengidentifikasi nilai amplitudo puncak tertinggi dalam data audio. Setelah itu, data audio akan disesuaikan sehingga nilai amplitudo puncak tersebut mencapai level yang telah ditentukan, misalnya 0 dB. Dalam flowchart peak normalisasi (Gambar 2), langkah-langkah tersebut dijelaskan secara visual. Dengan melakukan *peak normalization*, data audio dari berbagai sumber dapat diharmonisasi dan dibandingkan lebih mudah. Hasil dari peak normalization akan menghasilkan data yang lebih stabil dan memiliki rentang dinamis yang sesuai, memastikan bahwa data tersebut dapat diproses secara konsisten dalam analisis selanjutnya [3]. Dalam konteks penelitian ini, *peak normalization* menjadi langkah penting dalam mempersiapkan data sebelum dilakukan analisis lebih lanjut untuk menghasilkan hasil yang akurat dan informatif.



Gambar 2. Flowchart *peak normalization*

2.4 Tahap *Short-Time Fourier Transform* (STFT)

Langkah awal yang dilakukan adalah menganalisis spektral menggunakan *Short Time Fourier Transform* (STFT). STFT berfungsi untuk membagi sinyal audio menjadi jendela-jendela waktu kecil dan menghasilkan representasi frekuensi dari setiap jendela tersebut [4]. Sebelum dilakukan *transformasi Fourier* menggunakan algoritma FFT, sinyal audio dikenai pengaturan jendela seperti jendela Hamming atau Blackman. Hasil transformasi tersebut menyediakan informasi spektral dari sinyal audio dalam bentuk domain frekuensi.

Pada tahap ini, resolusi waktu dan frekuensi ditentukan oleh lebar jendela dan jumlah titik FFT yang digunakan. Dari proses STFT ini, dihasilkan *Spectrogram*, yaitu representasi visual dari perubahan spektrum frekuensi seiring waktu. *Spectrogram* memungkinkan pemantauan perubahan energi frekuensi pada sinyal audio sepanjang waktu.
Selanjutnya, analisis frekuensi pada setiap jendela waktu digunakan untuk mengidentifikasi nada atau *pitch* yang terdapat dalam sinyal audio. Metode deteksi puncak atau teknik pemrosesan sinyal lainnya dapat diterapkan untuk memperoleh informasi nada yang lebih akurat.

2.5 Konversi Ke MIDI

Dalam proses konversi audio ke MIDI, informasi frekuensi yang diperoleh dari analisis spektral menggunakan *Short Time Fourier Transform* (STFT) diubah menjadi data MIDI. Proses ini melibatkan beberapa langkah penting. Pertama, frekuensi dominan yang diidentifikasi dari setiap frame hasil STFT dikonversi menjadi nilai MIDI yang sesuai dengan skala musik yang telah ditentukan [5]. Konversi ini memerlukan penentuan hubungan antara frekuensi dan nilai MIDI, di mana frekuensi tinggi akan dikonversi ke catatan MIDI yang lebih tinggi, sementara frekuensi rendah akan dikonversi ke catatan MIDI yang lebih rendah.
Selain itu, durasi dari setiap catatan MIDI ditentukan berdasarkan durasi sinyal audio asli, di mana suara yang lebih panjang akan menghasilkan catatan MIDI yang lebih panjang. Intensitas suara juga mempengaruhi intensitas catatan MIDI yang dihasilkan, sehingga informasi ini juga diperhitungkan dalam proses konversi. Setelah nilai-nilai MIDI dihasilkan, langkah selanjutnya adalah membuat file MIDI dengan menggunakan data tersebut. File MIDI akan memuat informasi tentang catatan musik, seperti nota, durasi, dan intensitas dari sinyal audio yang telah diubah menjadi bentuk MIDI.

2.6 Desain Evaluasi Sistem

Dalam proses evaluasi sistem, akan dilakukan perhitungan untuk mendapatkan nilai akurasi akhir dari sistem yang telah dibuat. Tujuan dari evaluasi ini adalah untuk mengukur sejauh mana sistem mampu menghasilkan kesamaan melodi yang akurat. Evaluasi dilakukan dengan menguji sistem melalui perbandingan melodi original dengan hasil konversi MIDI yang dihasilkan oleh sistem. Kemudian, tingkat kesamaan melodi antara keduanya diukur. Setelah perbandingan dilakukan, langkah selanjutnya dalam evaluasi adalah menghitung akurasi. Akurasi digunakan sebagai metrik untuk menilai sejauh mana sistem berhasil menghasilkan kesamaan melodi yang sesuai dengan melodi referensi yang ada. Proses penghitungan akurasi melibatkan semua data audio yang telah diuji dengan menggunakan metode melody similarity untuk mendapatkan skor kesamaan. Dari hasil skor similarity tersebut, nilai akurasi dihitung dengan menggunakan rumus (2). Dalam rumus tersebut, skor hasil dari pengujian akan dijumlahkan dan dibagi dengan total data yang diuji, yang terdiri dari 11 dataset original dan data audio yang sudah dikonversi.

$$\text{Total Akurasi} = \frac{(MS1 + MS2 + \ldots + MSN)}{N} \quad (1)$$

Keterangan:
Total Akurasi = Total akurasi yang ingin dihitung.
MS = Nilai dari Melody Similarity
N = Jumlah data audio yang dievaluasi

3. Hasil Dan Pembahasan

Sistem dibangun harus cocok dengan lingkungan yang digunakan, termasuk sistem operasi Microsoft Windows 11 Home versi 64 bit pada laptop dengan spesifikasi CPU AMD Ryzen 7 5800H (Octa-core, hingga 4.4 GHz), RAM 16GB DDR4 3200MHz, dan Kartu Grafis AMD Radeon RX 6700M (6GB GDDR6 VRAM). Implementasi menggunakan bahasa pemrograman Python 3.1.0 dengan antarmuka berbasis HTML, CSS, dan JavaScript. Framework yang digunakan adalah Bootstrap 3 untuk frontend dan Flask 1.1 untuk backend, mempermudah penggunaan CSS dan Python.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

3.1 Antarmuka Sistem

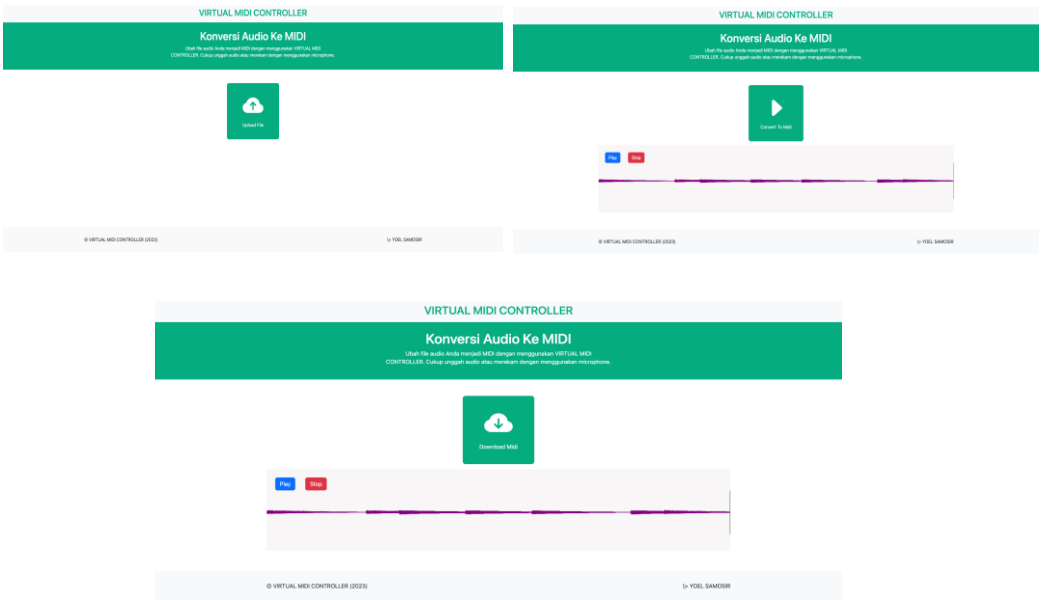
1. Tampilan Halaman Depan

Tampilan halaman depan menunjukkan tampilan awal yang muncul bagi pengguna ketika memulai sistem. Pada halaman depan ini, pengguna diberikan opsi untuk memilih metode konversi audio, yaitu melalui penggunaan microphone atau mengunggah file. Pada Gambar 3 merupakan tampilan halaman depan.

Gambar 3. Tampilan Halaman Depan

2. Tampilan Halaman Upload

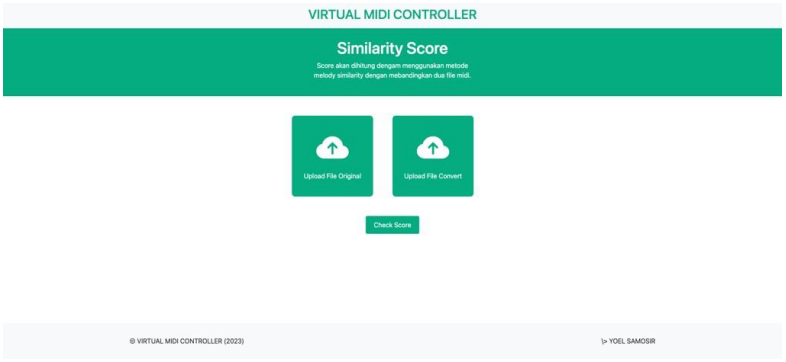
Tampilan halaman upload merupakan tampilan untuk mengkonversi audio ke Midi dengan menggunakan file upload. Setelah pengguna melakukan proses upload, Pengguna memiliki opsi untuk memutar audio yang telah diunggah dan juga dapat menghentikan pemutaran tersebut. Setelah mengunggah file, pengguna dapat melakukan konversi ke format MIDI dengan menekan tombol berwarna hijau. Hasil konversi dapat diunduh yang berupa file Midi. Pada Gambar 4 merupakan tampilan halaman Upload.



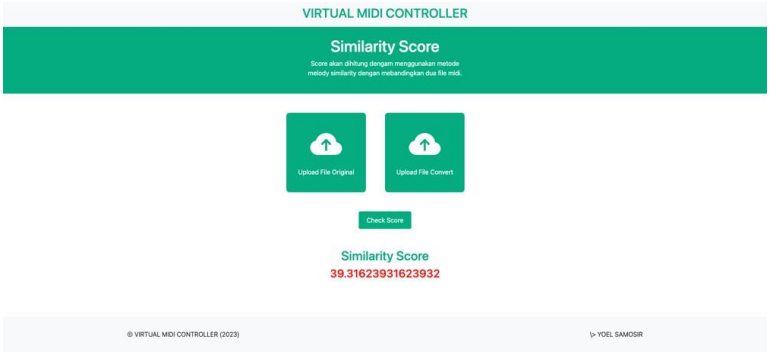
Gambar 4. Tampilan Halaman Upload

3. **Tampilan *Melody Similarity***

Tampilan *Melody Similarity* merupakan tampilan untuk melakukan *similarity score*. Pengguna diminta untuk mengunggah dua file, yaitu audio MIDI original dan audio MIDI hasil konversi menggunakan sistem yang dibuat. Setelah pengguna melakukan pengecekan kesamaan melody, hasil skor similarity akan ditampilkan dengan *font color* berwarna merah. Pada Gambar 5 dan 6 merupakan tampilan *melody similarity*.



Gambar 4. Tampilan Melody Similarity



Gambar 6. Tampilan Setelah Cek Similarity

3.2 Pengujian dan Evaluasi

Pada pengujian dan evaluasi memiliki tujuan utamanya adalah untuk memperoleh nilai akurasi dari total 11 data audio yang terdapat dalam dataset. Pengujian dilaksanakan dengan menggunakan metode *Melody Similarity* untuk menghitung skor kesamaan antara dua file MIDI yang dibandingkan. Dalam perbandingan tersebut, file MIDI pertama merupakan file asli dari dataset, sedangkan file MIDI kedua adalah hasil konversi menggunakan sistem yang telah dibuat. Pengujian ini menggunakan metode Kesamaan Melodi, dimana semakin tinggi nilai skor mendekati 100, maka kedua file MIDI tersebut semakin mirip. Sebaliknya, jika nilai skor mendekati 0, maka kedua file MIDI tersebut sangat berbeda. Hasil pengujian dengan metode Kesamaan Melodi dapat dilihat dalam Tabel 1.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

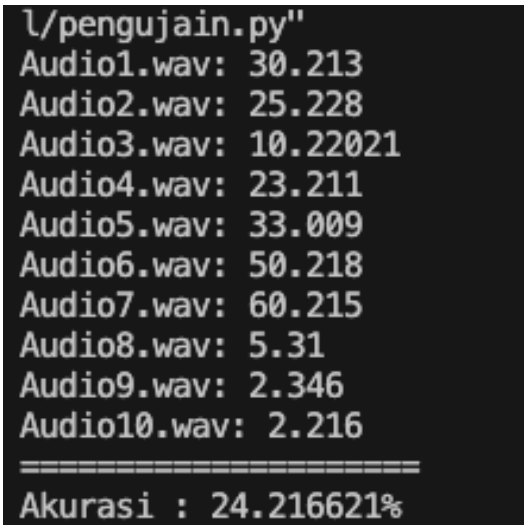
Tabel 1. Evaluasi Melody Similarity

Audio	*Melody Similarity*
Audio1.wav	30.213
Audio2.wav	25.228
Audio3.wav	10.22021
Audio5.wav	23.211
Audio6.wav	33.009
Audio7.wav	50.218
Audio8.wav	60.215
Audio9.wav	5.310
Audio10.wav	2.346
Audio11.wav	2.216

Setelah melakukan pengujian untuk mendapatkan skor kemiripan antara data audio asli dan data audio yang dibandingkan, selanjutnya dilakukan perhitungan akurasi dari total 11 data yang telah diuji menggunakan persamaan (1). Berikut adalah perhitungan total akurasi dari 11 data untuk mendapatkan akurasi akhir dari sistem yang dibuat:

Total Akurasi = (30.213 + 25.228 + 10.22021 + 23.211 + 33.009 + 50.218 + 60.215 + 5.310 + 2.346 + 2.21611) / 11
= 24.216621

Dari hasil perhitungan di atas, didapatkan bahwa akurasi sistem dengan total 11 data yang diuji adalah sebesar 24.216621%. Hasil ini menunjukkan bahwa akurasi sistem jauh dari 100%, yang mengindikasikan bahwa metode yang digunakan dalam proses konversi ke MIDI masih perlu diperbaiki. Gambar 7 menunjukkan hasil implementasi dari pengujian tersebut.



Gambar 7. Implementasi Pengujian

Dari penelitian yang telah dilakukan, kualitas audio yang baik memainkan peran penting dalam meningkatkan akurasi pendeteksian nada. Sinyal audio yang jernih dan jelas memungkinkan algoritma pendeteksian untuk lebih akurat mengidentifikasi frekuensi dan pola nada dalam audio tersebut. Sebaliknya, audio berkualitas rendah, seperti rekaman yang terdistorsi atau berisik, dapat menyebabkan informasi nada menjadi kabur dan mengakibatkan kesalahan dalam pendeteksian.
Selain itu, pemilihan parameter yang tepat dalam metode Short-Time Fourier Transform (STFT) juga mempengaruhi akurasi pendeteksian nada. STFT digunakan untuk menganalisis sinyal audio dalam domain frekuensi dengan membagi sinyal menjadi segmen-segmen waktu yang lebih pendek. Pengaturan parameter seperti panjang jendela dan overlap antar jendela dapat mempengaruhi resolusi frekuensi dan waktu dalam analisis STFT. Oleh karena itu, pemilihan parameter yang sesuai dengan karakteristik audio yang sedang diproses sangat penting. Pengaturan threshold yang optimal juga memiliki peran krusial dalam meningkatkan akurasi pendeteksian nada. Threshold digunakan untuk membedakan antara sinyal nada dengan sinyal kebisingan atau suara lainnya. Pengaturan threshold yang terlalu rendah dapat menyebabkan banyak sinyal kebisingan atau gangguan yang salah dianggap sebagai nada, sedangkan pengaturan threshold yang terlalu tinggi dapat menyebabkan nada yang lemah atau subtil tidak terdeteksi. Oleh karena itu, mencari nilai threshold yang tepat sangat penting untuk meningkatkan akurasi pendeteksian nada.

3. Kesimpulan

Berdasarkan hasil penelitian mengenai Konversi Suara ke MIDI Menggunakan *Short Time Fourier Transform*, dapat disimpulkan bahwa penggunaan STFT sebagai pendekatan untuk mengidentifikasi frekuensi suara dan mengkonversinya ke notasi MIDI dapat menghasilkan tingkat akurasi yang memadai, yaitu sebesar 24.216621%. Hasil penelitian mengidentifikasi bahwa faktor-faktor seperti kualitas audio, parameter STFT, dan pengaturan threshold memiliki pengaruh signifikan terhadap akurasi pendeteksian nada. Kualitas audio yang baik, pemilihan parameter STFT yang tepat, dan pengaturan threshold yang optimal dapat meningkatkan akurasi pendeteksian. Meskipun metode STFT memberikan hasil yang memadai, namun hasil konversi suara ke format MIDI masih memiliki keandalan yang kurang baik, dengan akurasi yang jauh dari 100%, yaitu sebesar 24.216621%. Dalam peningkatan konversi suara ke MIDI, perlu dipertimbangkan untuk mengoptimalkan faktor-faktor yang mempengaruhi akurasi sehingga informasi MIDI yang dihasilkan menjadi lebih andal dan mendekati akurasi yang lebih tinggi.

Referensi

1. Abdillah, F.N. 2017. Implementasi Algoritma Fast Fourier Transform (FFT) Dan Algoritma Harmonic Product Spectrum (HPS) Pada Tuner Gitar Berbasis Android. Universitas Kuningan. Kuningan.
2. Cadoz, C., & Wanderley, M.M. (2000). Gesture-Music. Proceedings of the International Computer Music Conference (ICMC).
3. Mubarok, A.B. Syauqy, D. Arwani, I. 2019. Sistem Pembacaan Nada Trumpet dengan Metode Fast Fourier Transform (FFT) Berbasis Embedded System. Universitas Brawijaya. Malang.
4. Mulyadi, Y. dan Daryana, H.A. 2020. DAW (Digital Audio Workstation) Technology In The Music Of West Java Traditional Theatre. Institut Seni Budaya Indonesia. Bandung.
5. Pratama, A.N. 2023. Pengembangan MIDI Controller Berbasis Microcontroller Dengan Mekanisme Sentuh. Universitas Negeri Yogyakarta. Yogyakarta