Laporan Pengayaan & Diskusi Analisis Fitur Audio

BAB I – PENDAHULUAN

Latar Belakang

Sinyal suara merupakan medium informasi yang kaya, tidak hanya memuat konten semantik (apa yang diucapkan), tetapi juga informasi paralinguistik dan non-linguistik seperti identitas pembicara, kondisi kesehatan, dan keadaan emosional. Ekstraksi fitur audio adalah fondasi kritis dalam disiplin ilmu *speech processing*, *music information retrieval* (MIR), hingga *clinical voice assessment*. Keberhasilan sistem cerdas yang memproses suara (misalnya, *Speech Emotion Recognition* atau SER) sangat bergantung pada kualitas representasi fitur yang diekstrak dari sinyal mentah. Laporan ini bertujuan untuk mengupas dan menganalisis berbagai aspek teknis dan metodologis dalam ekstraksi dan validasi fitur audio, berdasarkan kajian teoritis dan diskusi teknis mendalam.

Rumusan Masalah

Berdasarkan tinjauan metodologi dan diskusi teknis dalam pemrosesan sinyal audio, rumusan masalah yang akan dibahas dalam laporan ini adalah:

1. Bagaimana perbandingan prinsip, kelebihan, dan kekurangan antara metode ekstraksi fitur *Linear Predictive Coding* (LPC), *Mel-Frequency Cepstral Coefficients* (MFCC), dan *Perceptual Linear Prediction* (PLP)?
2. Bagaimana metodologi *Expert Weighted Evaluation* (EWE) dan statistik Cohen's Kappa digunakan untuk memastikan reliabilitas dan stabilitas *ground truth* pada data audio emosi atau kondisi klinis (batuk)?
3. Bagaimana penerapan fitur **CHROMA** berkontribusi dalam tugas-tugas *Music Information Retrieval* (MIR)?
4. Apa potensi dan tantangan penggunaan parameter mikro-prosodi seperti **Jitter** dan **Shimmer** sebagai biomarker klinis?
5. Bagaimana strategi yang efektif untuk integrasi fitur multi-modal dalam kerangka analisis suara yang umum?
6. Apa jawaban dan implikasi teknis terkait digitalisasi, *windowing*, *Audio Activity Detection* (AAD), kombinasi *Zero Crossing Rate* (ZCR) dan Energi, peran *Automatic Speech Recognition* (ASR) untuk analisis emosi, serta isu *brute-forcing* dalam pemilihan fitur?

Tujuan

Tujuan dari laporan ini adalah:

1. Menyajikan analisis komparatif mendalam mengenai metodologi ekstraksi fitur akustik utama.
2. Mengevaluasi peran teknik validasi data (Cohen's Kappa dan EWE) dalam meningkatkan reliabilitas dataset.
3. Membahas aplikasi dan relevansi fitur-fitur khusus (CHROMA, Jitter/Shimmer) di berbagai domain.
4. Menyajikan solusi teknis dan rekomendasi implementasi untuk isu-isu fundamental dalam *pipeline* pemrosesan audio digital.

BAB II – PEMBAHASAN

2.1 Perbandingan Metodologi Ekstraksi Fitur Spektral

Ekstraksi fitur adalah langkah pertama untuk mengonversi sinyal audio menjadi representasi numerik yang dapat diproses oleh algoritma *Machine Learning*.

Fitur	Prinsip Dasar	Kelebihan	Kelemahan
LPC	Memodelkan saluran vokal (traktus vokal) sebagai filter *all-pole* dengan mencari koefisien prediksi linier sinyal saat ini dari sampel sebelumnya.	Sangat efisien, komputasi ringan, ideal untuk aplikasi *real-time* dan perangkat terbatas.	Sensitif terhadap kebisingan (*noise*), kurang representatif terhadap persepsi pendengaran manusia.
MFCC	Meniru pendengaran manusia (skala Mel), menggunakan *Discrete Cosine Transform* (DCT) pada log energi filterbank Mel.	Standar emas di *speech recognition*, akurasi tinggi, robust terhadap variasi suara.	Berat secara komputasi (membutuhkan FFT, filterbank, Log, dan DCT), tidak ideal untuk perangkat sangat terbatas.
PLP	Menggabungkan	Menghasilkan	Lebih kompleks dan

	keunggulan LPC dengan prinsip persepsi (kurva kenyaringan dan *critical-band*).	representasi yang lebih natural dan mendekati respons auditori manusia. Lebih robust terhadap variasi *speaker*.	membutuhkan langkah pemrosesan tambahan dibandingkan LPC murni.

Analisis: Pemilihan fitur merupakan *trade-off* kritis. **LPC** dipilih untuk konteks *real-time* atau *resource-constrained*. **MFCC** adalah pilihan default untuk akurasi tertinggi pada lingkungan *offline* atau dengan *resource* memadai. **PLP** menjadi solusi kompromi yang menawarkan peningkatan kualitas perseptual dengan kompleksitas moderat.

2.2 Validasi dan Reliabilitas Data Audio

Kualitas label (emosi, kondisi batuk, dll.) pada data audio sangat penting.

Tantangan Reliabilitas

Tantangan utama berasal dari subjektivitas *annotator* (penilai) dan tingginya variabilitas akustik dalam kondisi rekaman yang berbeda-beda.

Hasil Eksperimen Validasi

- **Cohen's Kappa ():** Merupakan ukuran konsistensi antar-penilai yang lebih robust daripada akurasi sederhana karena memperhitungkan kesepakatan yang terjadi secara kebetulan. Hasil mengindikasikan ***Substantial Agreement*** (*Strong Agreement*) antar-penilai, membuktikan adanya konsensus kuat mengenai interpretasi emosi/kondisi.
- **Expert Weighted Evaluation (EWE):** Metodologi ini memberikan bobot lebih tinggi pada penilaian yang konsisten dengan *gold standard* (label kebenaran yang sudah terverifikasi) atau konsensus mayoritas. Hasil EWE dengan akurasi terhadap *gold standard* menunjukkan bahwa proses konsensus *ground truth* yang dihasilkan sangat stabil dan memiliki bias individu yang rendah.

Analisis: Kombinasi validasi menggunakan **Cohen's Kappa** dan **EWE** secara efektif mengurangi bias subjektif dan variabilitas. Proses ini memastikan bahwa dataset yang digunakan untuk pelatihan model memiliki label yang sangat reliabel, sehingga meningkatkan potensi kinerja dan generalisasi model *Machine Learning*.

2.3 Aplikasi Fitur CHROMA dalam MIR

CHROMA (*Chromagram* atau *Pitch Class Profile*) adalah fitur spektral domain musik yang merepresentasikan distribusi energi sinyal di 12 kelas nada (C, C#, D, ..., B), tanpa memandang oktaf.

Implementasi

CHROMA adalah tulang punggung untuk:

- **Chord Recognition:** Mengidentifikasi akor yang dimainkan pada interval waktu tertentu.
- **Key Detection:** Menentukan tonalitas (kunci) dari sebuah lagu.
- **Cover Song Identification:** Membandingkan kesamaan harmonis antara dua lagu yang dimainkan dengan oktaf atau *instrumentasi* berbeda.

Kelebihan dan Kelemahan

- **Kelebihan: Invarian terhadap Oktaf.** Properti ini membuat pola harmoni stabil meskipun dimainkan di register yang berbeda, menjadikannya fitur ideal untuk analisis harmoni.
- **Kelemahan: Kehilangan Detail Spektral.** Karena semua energi dari sebuah *pitch class* (misalnya, semua nada C dari semua oktaf) digabungkan menjadi satu, fitur CHROMA kehilangan informasi penting mengenai register vokal atau *timbre* instrumen.

Analisis: CHROMA dapat diimplementasikan dalam berbagai representasi: *scale-based* (untuk deteksi kunci), *chord-based* (untuk identifikasi akor), atau *PTR-based* (*Pitch Class Temporal Regularity*) untuk pola spesifik. Fitur ini sangat efektif, asalkan informasi register/timbre yang hilang tidak relevan dengan tugas yang diemban.

2.4 Mikro-Prosodi sebagai Biomarker Klinis

Mikro-prosodi merujuk pada fluktuasi minor dan cepat dalam sinyal vokal, yang dapat menjadi indikator sensitif bagi kondisi neurologis atau gangguan vokal (disfonia).

Jitter dan Shimmer

- **Jitter:** Fluktuasi kecil (*micro-variation*) dari periode fundamental sinyal suara ().
- **Shimmer:** Fluktuasi amplitudo dari periode fundamental sinyal suara.

Potensi Klinis

Fluktuasi yang meningkat (Jitter/Shimmer tinggi) seringkali mengindikasikan ketidakstabilan otot vokal, yang dapat menjadi gejala awal dari penyakit neurologis seperti Parkinson, Huntington, atau gangguan laring.

Analisis: Jitter dan Shimmer memiliki potensi tinggi untuk dijadikan **biomarker** digital karena dapat diekstraksi secara non-invasif. Integrasi fitur ini dengan fitur akustik tradisional (MFCC, LRR) dapat meningkatkan akurasi dalam klasifikasi klinis. Namun, terdapat tantangan signifikan terkait variasi individu dan perlunya **standarisasi protokol pengukuran** untuk memastikan interpretasi klinis yang hati-hati dan konsisten di berbagai laboratorium.

2.5 Integrasi Multi-Modal Features

Integrasi berbagai jenis fitur dari domain yang berbeda-beda (*multi-modal*) adalah arah utama dalam penelitian analisis suara modern.

Pendekatan Integrasi

Kombinasi yang efektif meliputi:

1. **Statistical Spectral Features:** (Contoh: *Spectral Centroid, Spectral Spread, Spectral Kurtosis, Skewness*).
2. **Domain-Specific Features:** (Contoh: **CHROMA** untuk musik, **MPEG-7** untuk kualitas audio, parameter **Prosodi** seperti Pitch, Energi, Durasi).

Kelebihan, Risiko, dan Solusi

- **Kelebihan:** Sistem yang dihasilkan menjadi lebih kaya informasi, generalis, dan robust terhadap variasi masukan.
- **Risiko:** Peningkatan dimensi fitur dapat menyebabkan *overfitting* (terutama pada dataset kecil) dan berpotensi memunculkan konflik di mana fitur-fitur yang berbeda memberikan informasi yang kontradiktif.
- **Solusi:** Harus diikuti dengan teknik **Feature Selection** (*mRMR: minimum Redundancy Maximum Relevance*) atau **Dimensionality Reduction** (*PCA, LDA*), serta **normalisasi** yang konsisten di semua domain fitur.

Analisis: Integrasi multi-modal adalah kebutuhan, bukan hanya pilihan. Ini memungkinkan sistem untuk menangkap aspek konten (ASR), kualitas (Jitter/Shimmer), dan emosi/harmoni (Prosodi/CHROMA) secara simultan, memberikan gambaran komprehensif tentang sinyal suara.

BAB III – DISKUSI TEKNIS FUNDAMENTAL

Bagian ini membahas enam aspek teknis penting yang membentuk dasar dari *pipeline* pemrosesan sinyal audio digital.

Digitalisasi Audio Analog

Digitalisasi adalah proses mengonversi sinyal audio kontinu (analog) menjadi representasi diskrit (digital). Proses ini krusial untuk memungkinkan pemrosesan komputasi. Proses digitalisasi melibatkan:

- **Sampling:** Penentuan laju sampel (). Menurut Teorema Nyquist, .
- **Kuantisasi:** Penentuan *bit depth* () untuk merepresentasikan amplitudo. Kuantisasi menentukan *Signal-to-Noise Ratio* (SNR) kuantisasi: .

Implikasi Teknis: Proses digitalisasi membuat data audio **robust terhadap kebisingan** yang sering dialami oleh sinyal analog selama transmisi atau penyimpanan.

Windowing Sinyal

Sinyal audio dibagi menjadi *frame* kecil untuk analisis frekuensi (melalui FFT). Namun, memotong sinyal secara tiba-tiba (Frame Rectangular) menciptakan diskontinuitas yang

menghasilkan ***Spectral Leakage***.

Fungsi Windowing (seperti Hamming dan Hanning) adalah fungsi bobot yang diaplikasikan pada setiap *frame* untuk membawa amplitudo di kedua ujung *frame* mendekati nol secara mulus.

Trade-off: *Rectangular window* menghasilkan resolusi frekuensi yang sangat baik tetapi *leakage* tinggi. *Hamming/Hanning* mengurangi *leakage* tetapi sedikit mengorbankan resolusi frekuensi.

Audio Activity Detection (AAD)

AAD adalah proses segmentasi suara dari non-suara (kebisingan atau keheningan). Pendekatan sederhana yang hanya berbasis *energi* dan *thresholding* seringkali gagal di lingkungan bising yang realistis.

Solusi: Perlu beralih dari metode sederhana ke algoritma berbasis *Machine Learning* (misalnya, *Hidden Markov Model* atau GMM) atau pendekatan statistik yang lebih canggih, terutama ketika rasio *Signal-to-Noise* (SNR) rendah.

Zero Crossing Rate (ZCR) dan Energi

ZCR adalah frekuensi sinyal berubah tanda (persilangan nol). **Energi** adalah intensitas amplitudo sinyal. Kedua fitur ini saling melengkapi:

- **Voiced Speech: Energi Tinggi** dan **ZCR Rendah** (contoh: vokal).
- **Unvoiced Speech / Fricatives: Energi Rendah** dan **ZCR Tinggi** (contoh: konsonan /s/, /f/).
- **Silence / Noise: Energi Rendah** dan **ZCR Bervariasi**.

Penerapan: Kombinasi ZCR dan Energi sangat efektif untuk *Voice Activity Detection* (VAD) awal dan segmentasi ucapan yang kasar.

Automatic Speech Recognition (ASR) untuk Analisis Emosi

ASR digunakan untuk mentranskripsikan konten semantik (kata-kata) dari sinyal suara. Dalam konteks analisis emosi:

- **Peningkatan Akurasi:** Analisis emosi menjadi lebih akurat ketika konten semantik diintegrasikan dengan fitur akustik (prosodi dan kualitas suara). Contoh: Kata "buruk" yang diucapkan dengan nada gembira mungkin menunjukkan sarkasme, yang hanya dapat dianalisis melalui integrasi teks.
- **Integrasi:** Pendekatan *multi-modal* menggabungkan representasi tekstual (misalnya, *embedding* dari model bahasa) dan representasi akustik (MFCC, Prosodi) sebelum klasifikasi.

Isu Brute-Forcing Features

Menggunakan secara "brute-force" semua fitur yang tersedia dalam jumlah besar (*high-dimensional feature vector*) tanpa penyaringan memiliki risiko:

1. **Overfitting:** Model cenderung menghafal data pelatihan dan gagal bergeneralisasi.
2. **Beban Komputasi:** Meningkatkan waktu pelatihan dan inferensi secara eksponensial.
3. **Curse of Dimensionality:** Kepadatan data menurun drastis seiring bertambahnya dimensi, membuat model sulit menemukan korelasi yang signifikan.

Solusi: Feature Selection (seperti mRMR) untuk memilih subset fitur yang paling informatif, atau **Dimensionality Reduction** (PCA/LDA) untuk memproyeksikan fitur ke ruang dimensi rendah tanpa kehilangan informasi yang signifikan.

BAB IV – KESIMPULAN

Berdasarkan analisis dan diskusi teknis yang telah dilakukan, dapat ditarik beberapa kesimpulan utama mengenai pemrosesan dan analisis fitur audio modern:

1. **Metodologi Ekstraksi Fitur:** Pemilihan antara LPC, MFCC, dan PLP bergantung pada **konteks aplikasi** (*trade-off* antara efisiensi *real-time* LPC, akurasi tinggi MFCC, dan kualitas perseptual PLP). Tidak ada satu fitur pun yang unggul di semua skenario.
2. **Validasi Data:** Validasi *ground truth* label audio adalah langkah wajib. Penggunaan **Cohen's Kappa** () dan **EWE** (Akurasi) secara signifikan meningkatkan **reliabilitas** dataset dengan memastikan kesepakatan penilai yang substansial.
3. **Aplikasi Spesifik:** Fitur **CHROMA** membuktikan kegunaannya yang unik dalam *Music Information Retrieval* berkat sifatnya yang invarian terhadap oktaf, meskipun ada pengorbanan informasi *timbre* atau register.
4. **Biomarker Vokal:** Parameter mikro-prosodi seperti **Jitter** dan **Shimmer** memiliki potensi yang layak untuk dijadikan **biomarker** klinis untuk deteksi dini gangguan vokal dan neurologis. Implementasi praktisnya memerlukan standardisasi pengukuran yang ketat.
5. **Arah Masa Depan: Integrasi Multi-Modal** (statistik spektral, prosodi, dan domain-spesifik) merupakan arah penelitian kontemporer yang menjanjikan, karena menghasilkan sistem analisis suara yang lebih kaya, komprehensif, dan **robust**.
6. **Fondasi Teknis:** Aspek fundamental seperti digitalisasi yang tepat, penggunaan *windowing* yang meminimalkan *spectral leakage*, deteksi aktivitas suara yang cerdas (*AAD/ZCR/Energi*), dan **Feature Selection** yang hati-hati adalah fondasi yang wajib dikuasai untuk membangun *pipeline* analisis audio modern yang sukses.