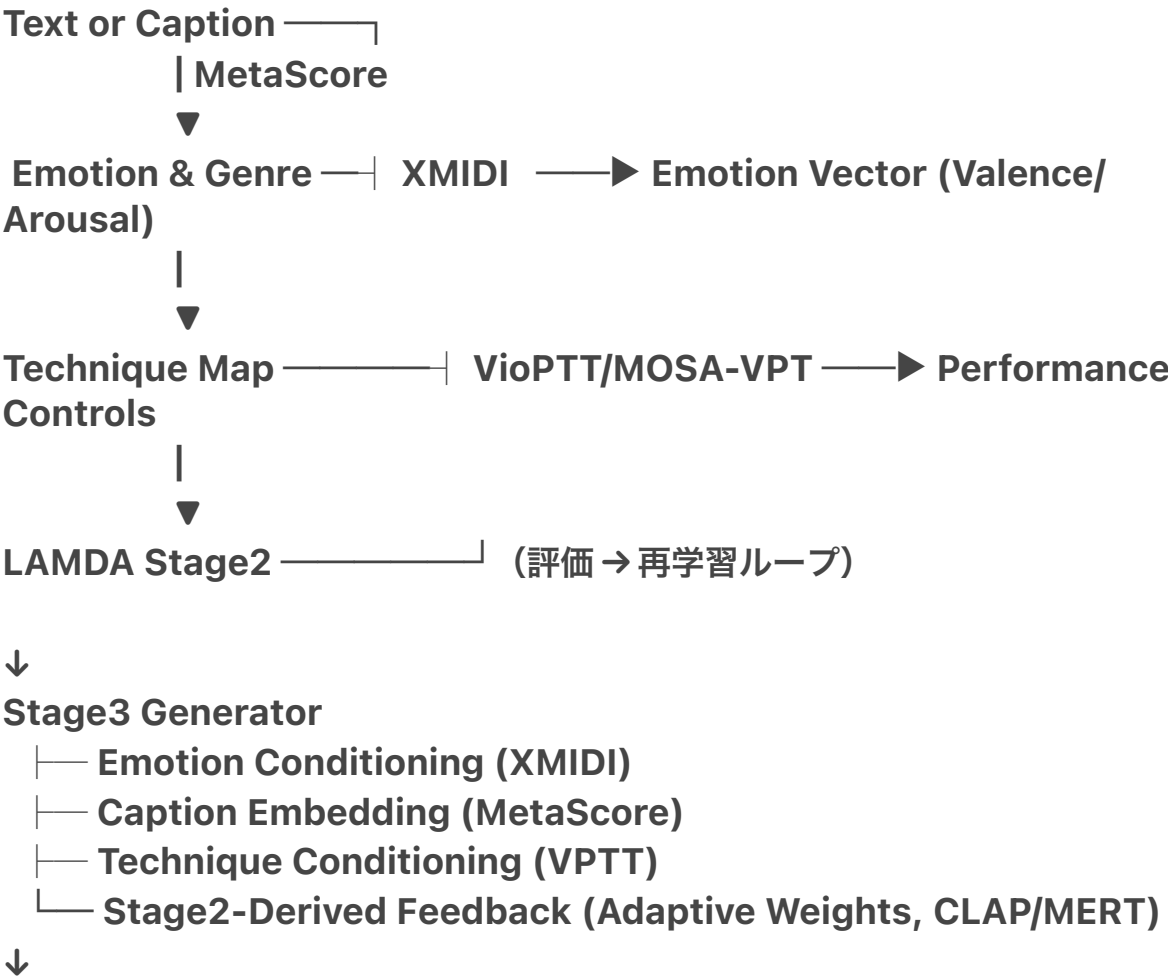


🌟 Stage3 ロードマップ (XMIDI・VioPTT/MOSA-VPT・MetaScore 統合)

1. フェーズ概要

| 項目 | 説明 |
|--------|---|
| フェーズ名 | Stage3: Context-Aware Composition Pipeline |
| 目的 | Stage2 で得た「評価知能」を活用し、音楽生成 (MIDI 出力) ヘフィードバックする |
| 技術中核 | XMIDI (感情・ジャンル教師) + VioPTT/MOSA-VPT (奏法モデル) + MetaScore (テキストキャプション生成) |
| 出力形式 | Multi-Track MIDI (Emotion/Genre/Technique/Caption タグ付き) + Loop Summary + Auto Score |
| 成功 KPI | 再現性 ↑ (同属性で近似曲生成) / 創造性維持 / スコア p50≥75, p90≥88 |

2. 構造図 (論理パイプライン)



MIDI + Captions + Metrics

3. 実装ステップ

| フェーズ | 実施内容 | 成果物 |
|--|---|---|
| Step 1 : XMIDI 感情・ジャンル教師導入 | labels_schema.yaml を XMIDI 語彙に統一。 assign_labels.py で valence/arousal, genre を自動推定。 | loop_summary.csv に label.emotion, label.genre 列を追加。 |
| Step 2 : MetaScore キャプション生成 | scripts/generate_music_captions.py を作成。Stage2 のメトリクス+歌詞/楽曲情報を LLM に入力。 | label.caption 列を生成し、Emotion/Genre と連動。 |
| Step 3 : VioPTT/MOSA-VPT 奏法制御導入 | configs/labels/technique_map.yaml で奏法ラベルを定義。 batch_articulation_renderer.py で奏法合成。 | 合成データセット (奏法別 MIDI) + 奏法マップ (Key/CC 対応表)。 |
| Step 4 : Stage3 Generator 開発 | ml/stage3_generator.py : XMIDI・MetaScore・VPTT を条件トークンとして学習。 | 感情・テキスト・奏法条件付きの自動作曲モデル。 |
| Step 5 : LAMDA Feedback 統合 | Stage2 のメトリクスを Stage3 学習に報酬として取り込み (RL or weighting)。 | 自動評価・改善ループ (自己学習的生成)。 |

4. 出力・レポート

| 出力物 | 説明 |
|---------------------------|---|
| stage3_summary.json | Stage3 生成の統計 (Emotion, Caption, Technique 別性能)。 |
| loop_summary_stage3.csv | 生成結果のメタ情報 (感情・ジャンル・奏法・スコア)。 |
| music_captions.log | MetaScore 生成ログ (要約・情景文)。 |
| emotion_distribution.json | 感情ベクトルの分布ヒートマップ。 |
| stage3_ab_report.md | A/B 比較レポート (生成品質・評価スコア)。 |

5. 成功指標 (KPI)

| 軸 | 指標 | 目標値 |
|---|----|-----|
|---|----|-----|

| | | |
|--------|-----------------------------------|-----------|
| 品質 | Stage2再評価後の p50/ p90 | ≥75 / ≥88 |
| 文脈一致 | CLAP/MERT 類似度 (text_audio_cos) | ≥0.65 |
| 感情再現率 | Emotion分類精度 (XMIDI) | ≥85% |
| 奏法再現率 | Technique ラベル一致率 | ≥80% |
| 創造的多様性 | 曲間平均類似度 | ≤0.55 |
| 再現性 | 同条件再生成のスコア分散 | ≤5% |

6. 今後の展望

- **MetaScore 2.0**：音楽構造を「物語の章」として説明できるキャプションを自動生成。
- **VPTT 拡張**：他楽器 (Strings, Piano, Winds) への奏法転移学習。
- **XMIDI Fine-Tune**：自作データで valence/arousal の再学習、より日本語感情語彙へローカライズ。
- **Stage4 Preview**：自動リライト (「感情 → 旋律 → 調整 → 再評価」の完全生成ループ)。

Stage3 ロードマップ では、拡張パック群 (CLAP/MERT・EMOPIA・MT3+ASAP・MuseCoco・LP-MusicCaps) を以下のように体系的に統合しています。

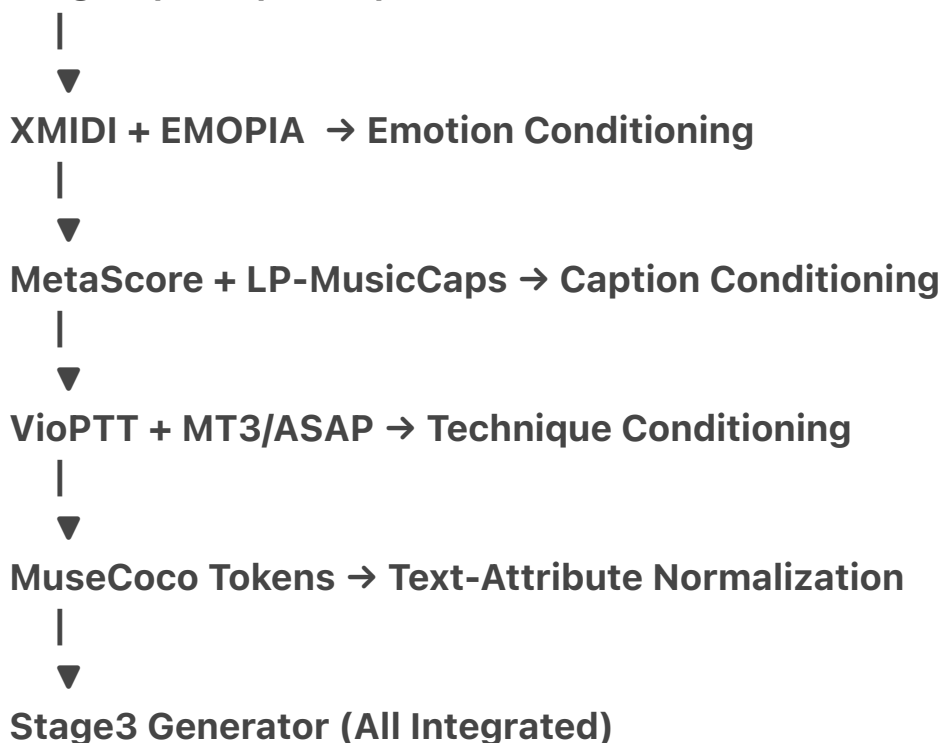
拡張パックとの整合 (現ロードマップ内での位置づけ)

| 拡張パック | 組み込み位置 | 目的・役割 | ロードマップ参照 |
|------------------------------|---------------------|--|--|
| CLAP / MERT | Stage2→Stage3 全体 | 音声-テキスト対応の基盤 (text_audio_cos・ audio_context) | Stage2 完結済 → Stage3 Feedback 項目に 継承 |
| EMOPIA | Step1：XMIDI 感情教師 | Valence/Arousal の正規化軸として Emotion 空間を補強 | XMIDI → Emotion Vector 出力部に統合 |
| MT3 + ASAP/ nASAP | Stage2→VioPTT 間 | 転写・演奏表現モデルとして Humanizer/ Performance 学習を補強 | Step3：VioPTT 奏法制御導入後、 テンポゆらぎ補正 へ反映 |

| | | | |
|---------------------|------------------|--|--------------------------------------|
| MuseCoco | Stage3 Generator | キャプションを属性トークン化 ([genre] [mood]) し生成安定性を高める | Step4 : Stage3 Generator 条件入力構造に追加予定 |
| LP-MusicCaps | MetaScore 強化 | LLM 生成キャプションを近傍補完して語彙カバレッジを拡大 | Step2 : MetaScore キャプション生成の補助学習 |

拡張パックの流れと統合位置イメージ

Stage2 (CLAP/MERT)



計画上の扱い

- **CLAP/MERT** はすでに実装済み (Stage2 完結) で、Stage3 では「報酬フィードバック入力」として活用。
- **EMOPIA / XMIDI** は「感情・ジャンル教師フェーズ (Step1)」に統合。
- **MT3 + ASAP/nASAP** は「演奏表現・Rubato 学習」のデータパイプとして Stage2 から接続。
- **MuseCoco / LP-MusicCaps** は「MetaScore 強化フェーズ (Step2-4)」に吸収。

つまり拡張パックは、単なる“オプション”ではなく、
Stage3 ロードマップの中核構成要素 として組み込まれています。