

# 音声の感情および意味内容を植物形態へ動的に反映するプロ シージャル生成手法 – LLMによる文脈解析とフィロタキシスを用いた構造 決定 –

滑川 裕里瑛<sup>1,a)</sup> 中村 亮太<sup>1,b)</sup>

受付日 2025年12月19日, 採録日 2025年12月19日

**概要:** 本研究は, 音声の感情価と意味内容をリアルタイムで解析し, その高次情報を植物学的規則に基づく 3D モデル生成へ動的にマッピングする手法を提案する. LLM を用いて抽出した感情と語彙 (意味内容) を, 花卉の色彩・鋭度といった連続値パラメータに加え, フィロタキシス (葉序) や花序といった構造規則の切り替えに適用する点が新規である. これにより, 自然律という幾何学的制約のもと, 美的整合性を保ちつつ, 発話者の言語化困難な内面を直感的に表現できるインタフェースを実現する.

**キーワード:** 音声解析, LLM, プロシージャル生成, フィロタキシス (葉序), 感情表現

YURIE NAMEKAWA<sup>1,a)</sup> RYOTA NAKAMURA<sup>1,b)</sup>

Received: December 19, 2025, Accepted: December 19, 2025

## 1. はじめに

近年, 音響特徴量を用いた映像生成は広く行われているが, 発話者の意図や感情といった高次情報を形状へ反映する手法は十分ではなかった. 本研究では, 音声の「感情価」と「意味内容」をリアルタイムに解析し, 植物学的規則に基づく 3D モデルの生成パラメータへ動的にマッピングする手法を提案する. 本手法は, LLM を用いて入力音声から感情および形状決定因子となる語彙を抽出し, これを花卉の色彩・鋭度といった連続値パラメータだけでなく, フィロタキシス (葉序) や花序といった構造規則の切り替えに

適用する点に新規性がある. これにより, 不定形になりがちな感情の視覚化に対し, 自然律という幾何学的制約を与えることで美的整合性を保ちつつ, 言語化困難な内面を直感的かつ豊かに表現可能なインタフェースを実現する.

## 2. 関連研究

### 2.1 音声解析と感情ベースの映像生成

音声から感情を抽出し, それを視覚的表現に変換する研究は, 主に顔表情生成や抽象的なビジュアルライザの分野で進展している.

- MEAD (Multi-view Emotional Audio-visual Dataset) [1]: 音声を感情 (Emotion) と発話内容 (Content) に分離し, 感情の強度に応じた顔表情を生成するモデルが提案されている.
- PAD モデルを用いた音声映像合成 [2]: Pleasure (快楽), Arousal (覚醒), Dominance (支配) の 3 軸を

<sup>1</sup> 情報処理学会  
IPSJ, Chiyoda, Tokyo 101-0062, Japan

<sup>†1</sup> 現在, 武蔵野大学データサイエンス学部データサイエンス学科  
Presently with Musashino University

<sup>a)</sup> lucky854854@gmail.com

<sup>b)</sup> ryonaka@musashino-u.ac.jp

用いて、音声信号から動的に感情表現を生成する手法が存在する。これらは人型モデルや単純な物理量へのマッピングに限定されており、複雑な幾何学的構造（植物形態など）への文脈的な反映には至っていない。

## 2.2 LLM によるプロシージャル 3D モデリング

近年、自然言語による記述（プロンプト）から、プロシージャルな 3D モデルを生成する手法が注目されている。

- 3D-GPT [3]: LLM を用いてテキスト記述を解析し、Procedural Content Generation (PCG) の関数呼び出しへと変換することで、編集可能な 3D シーンを構築するフレームワークが提案されている。
- ShapeCraft [4]: 文脈を構造化されたサブタスク（グラフ）に分解し、再帰的に 3D アセットを構築する手法が示されている。これらの研究は静的なテキスト入力を前提としており、リアルタイム音声の「感情の揺らぎ」を動的な構造変化に結びつけた例は少ない。

## 2.3 植物形態の数理モデルと音響連動

植物の幾何学的規則（フィロタキシスや L-system）を用いた生成手法は、自然界の美的整合性を保つための強力な制約として機能する。

- フィロタキシスの数理的頑健性 [5]: 黄金角に基づく葉序パターンが、ノイズや外部刺激に対してどのように自己組織化されるかの研究が行われている。
- Unity を用いたプロシージャル・オーディオ・ビジュアライゼーション [6]: 音響の周波数成分（スペクトラムデータ）をフィロタキシスのパラメータに直接マッピングし、リアルタイムに形状を変化させる実装例が報告されている。

本研究は、これら低次の音響連動に対し、LLM による「意味・文脈」の解釈を介在させることで、より意図的かつ表現豊かな形態変容を実現する。

## 3. 既存手法

従来の音声可視化およびインタラクティブな形態生成手法は、大きく分けて以下の 2 つのアプローチに分類される。

### 3.1 音響物理量に基づく動的生成

従来のリアルタイム・ビジュアライザの多くは、音声信号から抽出される音響特徴量（振幅、周波数、スペクトラムデータ等）を直接的に形状パラメータへマッピングする手法を採用している [5], [6]。

- 特徴: 音の大きさに応じた物体の膨張や、ピッチの変化に伴う色彩の変容など、物理的な音の変化に対して高い追従性を持つ。
- 限界: これらの手法は「音の物理現象」の可視化には適しているが、発話内容に含まれる「文脈（コンテキ

スト）」や「意図」、あるいは「複雑な感情の機微」を形状に反映することはできない。その結果、生成される映像は抽象的かつパターン化されたものに留まる傾向がある。

### 3.2 テキスト解析に基づく静的生成

自然言語処理（NLP）を用いた形態生成の研究では、入力されたテキストの意味内容から 3D モデルを構築する手法が提案されている [3], [4]。

- 特徴: 「鋭い花」「暗い森」といった具体的な語彙から、意味的に整合性の取れた形状やカラーパレットを生成することが可能である。
- 限界: これらの手法は、あらかじめ用意された完成済みのテキストを解析することを前提としており、音声入力のように刻々と変化する「時間的連続性」や、声のトーンに含まれる「非言語的な感情情報」をリアルタイムに統合して形態を更新する設計にはなっていない。

### 3.3 本研究の立ち位置

本研究の提案手法は、これら従来手法の欠点を補完するものである。LLM を用いることで、既存の音響解析では到達できなかった「意味的・感情的文脈」の抽出を行い、それを植物学的な構造規則（フィロタキシス等）にマッピングする。これにより、物理的な音の動きへの追従性と、高次文脈情報の反映を両立させ、不定形な感情を美的整合性の高い植物形態へと昇華させる点に独自性がある。

## 4. 提案手法

本手法は、LLM による高次情報抽出と、それに基づくプロシージャルな構造決定の 2 段階で構成される。

### 4.1 LLM による文脈・感情解析

音声認識により得られたテキストを LLM に入力し、以下の情報をリアルタイムで抽出する。

- 感情価 (Valence): 発話のポジティブ・ネガティブの度合いを数値化し、色彩の基本トーンを決定する。
- 形状決定語彙: 「鋭い」「柔らかい」「広がる」といった形態のメタファーを持つ語彙を抽出し、幾何学的パラメータへ変換する。

### 4.2 構造規則の切り替えとパラメータマッピング

抽出された情報は、単なる連続値の変化だけでなく、植物の「構造規則（アルゴリズム）」そのものの切り替えに適用される。

- フィロタキシス（葉序）の制御: 感情の安定度に応じて、黄金角等の発散角を変化させ、密集度や規則性を動的に更新する。

- 花序パターンの選択: 文脈の複雑度に基づき, 単頂花序や無限花序といった分岐構造のアルゴリズムを選択する.

## 5. 実験

提案手法を実装したインタラクティブ・システムを構築した. ユーザーがマイクを通じて自由に発話を行い, その内容がリアルタイムで 3D の植物形態 (花や葉の配列) へと変換されるプロセスを観察した.

## 6. 結果と考察

生成された 3D モデルは, 入力音声の感情的な起伏と文脈上の意味を, 視覚的な形状変化として一貫性を持って表現することが確認された. 特に, 音響特徴量のみを用いた手法と比較して, 言葉の意味 (「鋭い」という言葉に反応して花卉が尖る等) が直接的に反映されることで, 制作者の意図がより明確に視覚化されることが示唆された. 自然律に基づく幾何学的制約により, ランダムな形状変化よりも美的整合性が高く, ユーザーにとって直感的で豊かな表現が可能となった.

## 7. おわりに

本研究では, LLM による文脈解析を植物学的規則に接続することで, 音声の高次情報を動的に反映するプロシージャル生成手法を提案した. 今後は, より多様な植物学的メタファーの導入や, 空間音響を介した多角的な没入体験の提供について検討する.

## 参考文献

- [1] Wang, Y., et al.: MEAD: A Large-scale Audio-visual Dataset for Emotional Talking-face Generation, *ECCV*, 2020.
- [2] Zhang, J., et al.: Emotional Audio-Visual Speech Synthesis Based on PAD, *IEEE*, 2024.
- [3] Sun, T., et al.: 3D-GPT: Procedural 3D Modeling with Large Language Models, *NeurIPS*, 2023.
- [4] Ye, J., et al.: ShapeCraft: Decomposing Complex Natural Language into Structured 3D Tasks, *arXiv*, 2024.
- [5] Mirabet, V., et al.: Noise and Robustness in Phyllotaxis, *PLOS Computational Biology*, 2012.
- [6] AksAman: Phyllotaxis: A procedural audio visualization project in Unity, *GitHub Repository*, 2023.