Music as Programs: Towards Understanding Musical Structure via Program Induction

Robie Gonzales, Marta Kryven
Dalhousie University

Motivation

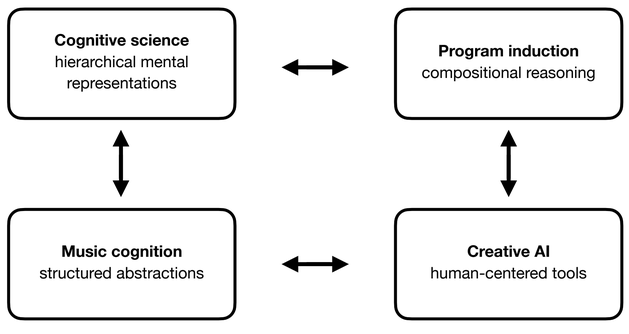
- Music is highly structured, reflecting both cultural traditions (shared transformations) and cognitive constraints (mental representations)
- Program induction models compositional concept learning, a natural fit for music's generative structure
- LLMs as program proposers make program induction tractable
- A foundational step, testing whether LLMs can induce canonical musical transformations with structured scaffolds
- Towards interpretable, controllable, human-centered creative AI that reasons with music like humans do

Cognitive science + program induction

- A rich body of work in cognitive science demonstrates that human concept learning is fundamentally *compositional*: we build complex ideas by systematically combining simpler elements
- *Program induction:* the process of inferring executable programs that generate observed patterns
- Music is an ideal testbed due to its hierarchical, generative nature

LLMs as program proposers

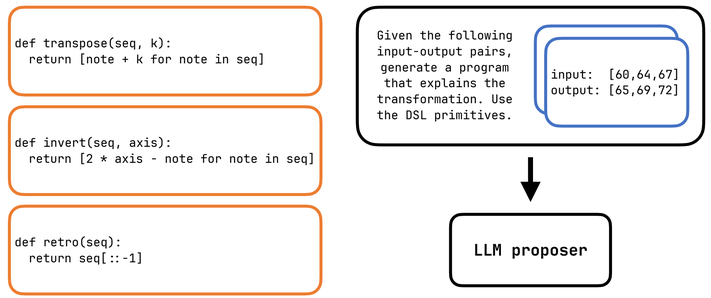
- A key challenge in program induction is the *curse of compositionality*: the space of possible programs grows exponentially, making exhaustive search intractable
- We leverage LLMs as stochastic program proposers to make search tractable. LLMs can:
  - Generate candidates in Turing-complete languages
  - Amortize search through learned priors over program structure
  - Incorporate context from input-output examples
  - Scale to realistic program complexity



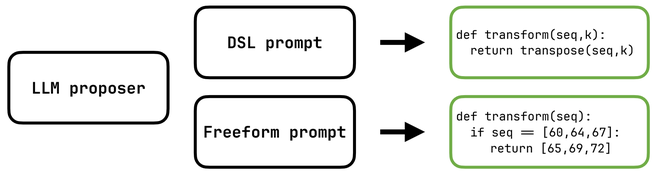
Research questions

1. Do LLMs learn generalizable transformation rules?
2. Can structured representational scaffolding, in the form of a DSL, bias LLMs towards genuine rule induction

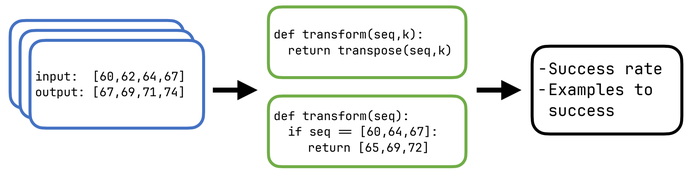
Approach



- We represent musical transformations as Python functions over MIDI pitch sequences. Each task follows a loop:
  1. Show the model a single input–output example
  2. Prompt it to propose a program explaining the mapping
  3. Execute the program on held-out probes
  4. If it fails, reveal an additional example and repeat



- Given input-output examples, an LLM proposes candidate programs using one of three representational scaffolds:
  - **DSL**: prompt provides primitive signatures and their Python implementations; solutions must call these primitives
  - **Vocab**: prompt lists primitive names; no implementations
  - **Freeform**: unconstrained Python generation



- At each sampling step, induced programs are tested on unseen sequences probing longer lengths, register shifts, and adversarial cases

Results

Transformations	Success Rate (%) ↑	Examples to Success ↓
Transpose (T_k)		
DSL	**97**	$\mathbf{2.0 \pm 0.6}$
Vocab	91	2.9 ± 1.1
Freeform	83	4.1 ± 1.5
Retrograde (R)		
DSL	**81**	$\mathbf{2.3 \pm 0.8}$
Vocab	68	3.2 ± 1.2
Freeform	37	4.8 ± 1.8
Inversion (I_{axis})		
DSL	**61**	$\mathbf{2.9 \pm 1.3}$
Vocab	27	4.3 ± 1.9
Freeform	12	5.2 ± 2.3

Table 1: LLMs learn musical transformation rules from just 2-3 examples when scaffolded with a DSL, achieving high success. Performance drops with weaker scaffolds (vocab-only or freeform).

- Structural scaffolds bias LLMs toward generalizable rule induction
- Program induction with LLMs is tractable in music with the right representation
- Mirrors cognitive findings, suggesting a computational probe of musical reasoning

Implications

- Creative AI: interpretable, controllable, human-centered tools
- Cognitive science: music as a testbed for program-like concept learning
- Foundational step: proof-of-concept for symbolic + neural hybrids in creative domains

Next steps

- Library learning: models that accumulate reusable motifs, transformations, and style grammars
- Bayesian approaches: refining candidate programs through probabilistic inference
- Beyond atomic ops: toward richer compositional hierarchies in real music corpora

Broader impact

- A paradigm for computational cognitive science in creative domains
- Grounding generative models in human-like concept learning — interpretable, compositional, falsifiable
- Towards interpretable, controllable, human-centered tools
- Empowers creators and researchers to co-create with AI that reasons in fundamentally human-like ways