REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
17-04-2020	Final Report	31-Jul-2015 - 30-Jan-2017

4. TITLE AND SUBTITLE
Final Report: MUSICA: MUSical Improvising Collaborative Agent

5a. CONTRACT NUMBER
W911NF-15-1-0460

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Arizona
Biomedical Engineering
Post Office Box 210158, Room 510
Tucson, AZ 85721 -0158

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
67833-CS-DRP.3

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Clayton Morrison
UU	UU	UU	UU		19b. TELEPHONE NUMBER 000-000-0000

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 10-Jun-2020

Agency Code:

Proposal Number: 67833CSDRP | **Agreement Number: W911NF-15-1-0460**

INVESTIGATOR(S):

Name: Clayton T. Morrison
Email: claytonm@email.arizona.edu
Phone Number: 0000000000
Principal: Y

Organization: **University of Arizona**
Address: Biomedical Engineering, Tucson, AZ 857210158
Country: USA

DUNS Number: 806345617 | EIN: 866004791
Report Date: 30-Apr-2017 | Date Received: 17-Apr-2020
Final Report for Period Beginning 31-Jul-2015 and Ending 30-Jan-2017
Title: MUSICA: MUSical Improvising Collaborative Agent
Begin Performance Period: 31-Jul-2015 | **End Performance Period:** 30-Jan-2017
Report Term: 0-Other
Submitted By: Clayton Morrison | Email: claytonm@email.arizona.edu
Phone: (000) 000-0000
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 1 **STEM Participants:** 1

Major Goals: The original goal of the MUSICA project is to develop a real-time Jazz improvisation system capable of interacting meaningfully with human jazz musicians -- this is referred to as the Trading Fours (T4) use case scenario within the Communicating with Computers Program. In order for this to be possible, MUSICA must be able to do the following:
(1) Take as input a digital music representation of music (whether originating from a quantization facility for real-time audio or from a representation of a score)
(2) Analyze the music excerpt to identify patterns recognizable by a human musician as significant
(3) Formulate a response the both (a) would be recognizable by a human musician as referencing musical patterns/features in the input score and/or from previous interactions while also (b) contributing novel musical structure.
(4) Output that response in both a score and audio form that a human musician can recognize/read.
Step (3) is the key step to achieving musical communication between a human and machine: MUSICA will need to demonstrate that it "hears" what the human musician is expressing while also contributing novel musical structure to the interaction.

In Mid 2016, a second use case for MUSICA was introduced: musical composition by conversation (CbC).
In composition by conversation, MUSICA interacts with human musicians through a natural language interface while the human and Musica collaborate on the shared task of creating a musical score. The goal is to develop the natural language interface capable of parsing text into a representation that MUSICA can then interpret as expressions about a score, with both the human and MUSICA contributing to edits made to the score.

Accomplishments: In service of the CbC use case, during this phase the team made significant progress in the design and implementation of a framework for representing Musical Elementary Composable Ideas (MusECI) in Python, heavily inspired by coPI Donya Quick's experience with the Kulita system.
The main progress toward CbC involved developing the Muser Musica module to interface with the Cogent dialogue system, implementing a facility to translate between TRIPS parser logical form and Musica actions and MusECI representations. This work was published in Quick & Morrison 2017.

In this phase, the team also made progress in the development of the MUSICA agent architecture. By the end of 2017, MUSICA consisted of the following modules:
Pattern Toolbox: a collection of algorithms including the fluent pattern finding framework, hidden Markov models, and the hierarchical Dirichlet Process hidden semi-Markov model.
Assumer: initial design concepts but not yet implemented - the Assumer will be responsible for reasoning about

background musical knowledge, to infer context and assumed shared musical conventions
Generator: Adaptation of Kulita musical score generation functions into Python
Perceiver: Start of MusicXML and MIDI import into MusECI data structures
Visualizer: Start of representations of MusECI scores in piano roll and MuseScore-rendered score layout
Performer: Start of export of MusECI to MusicXML and MIDI files
Muser: Interface of Musica to the Cogent Collaborative Problem Solving Dialogue System (includes wrapper to the TRIPS parser)

During this phase, progress was also made on automating the analysis of identification of musical patterns within MusicXML score representations, in support of the Trading Fours (T4) use case.
This included additional transcription of jazz solos, so that the transcription corpus now has over 150 MusicXML representations.
This phase we also expanded music pattern identification to search for novel repeated patterns. One way to identify novel repeated patterns is to use methods from data compression. This phase we adapted the SIATEC Music repeated pattern algorithm developed by David Meredith. Based on an analysis of SIATEC performance in identifying musically interesting patterns, we found that SIATECs compression optimization ends up identifying many small, spurious repeated motifs. To ameliorate this, we extended the SIATEC algorithm to incorporate a set of heuristic constraints and filters that then are hypothesized to be considered more musically interesting. A preliminary study this phase found that the augmented framework does a better job at identifying musically significant patterns than SIATEC alone, but a corpus annotated with patterns that jazz musicians deem as interesting or significant is needed for a quantitative evaluation.

Training Opportunities: The grant supported training for the following:
Paul Hein, 100% supported, 100% FTE summer, MS in Computer Science, May 2017-2019.
Supported under this grant, Paul Hein earned a BS in Computer Science in 2017.

Results Dissemination: Nothing to Report

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Co PD/PI
Participant: Clayton T Morrison
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Paul D Hein
Person Months Worked: 8.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Consultant
Participant: Christopher Herald

Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: PD/PI
Participant: Kelland Thomas
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co PD/PI
Participant: Donya Quick
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Co PD/PI
Participant: Ben Grosser
Person Months Worked: 15.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Proceedings of the Forty-Third International Computer Music Conference (ICMC)
Publication Identifier Type: Other Publication Identifier: http://arxiv.org/abs/1709.02076
Volume: Issue: First Page #:
Date Submitted: 4/16/20 12:00AM Date Published: 7/1/17 7:00AM
Publication Location:
Article Title: Composition by Conversation
Authors: Donya Quick, Clayton T. Morrison
Keywords: knowledge representation, music information retrieval, natural language processing
Abstract: Most musical programming languages are developed purely for coding virtual instruments or algorithmic compositions. Although there has been some work in the domain of musical query languages for music information retrieval, there has been little attempt to unify the principles of musical programming and query languages with cognitive and natural language processing models that would facilitate the activity of composition by conversation. We present a prototype framework, called MusECI, that merges these domains, permitting score-level algorithmic composition in a text editor while also supporting connectivity to existing natural language processing frameworks.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Proceedings of the Thirty-Fourth International Conference on Machine Learning (ICML)
Publication Identifier Type: Other Publication Identifier: https://arxiv.org/abs/1707.06756
Volume: Issue: First Page #:
Date Submitted: 4/17/20 12:00AM Date Published: 8/18/17 7:00AM
Publication Location:
Article Title: An Infinite Hidden Markov Model with Similarity-biased Transitions
Authors: Colin R. Dawson, Chaofan Huang, Clayton T. Morrison
Keywords: hidden Markov model, hierarchical Dirichlet process, time series, Markov jump process
Abstract: We describe a generalization of the Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) which is able to encode prior information that state transitions are more likely between "nearby" states. This is accomplished by defining a similarity function on the state space and scaling transition probabilities by pair-wise similarities, thereby inducing correlations among the transition distributions. We present an augmented data representation of the model as a Markov Jump Process in which: (1) some jump attempts fail, and (2) the probability of success is proportional to the similarity between the source and destination states. This augmentation restores conditional conjugacy and admits a simple Gibbs sampler. We evaluate the model and inference method on a speaker diarization task and a "harmonic parsing" task using four-part chorale data, as well as on several synthetic datasets, achieving favorable comparisons to existing models.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Communicating with Computers
Interim Progress Report

(1) Submissions or publications under ARO sponsorship **during this reporting period.** List the title of each and give the total number for each of the following categories:

(a) Papers published in peer-reviewed journals

(b) Papers published in non-peer-reviewed journals

(c) Presentations

i. Presentations at meetings, but not published in Conference Proceedings

ii. Non-Peer-Reviewed Conference Proceeding publications (other than abstracts)

iii. Peer-Reviewed Conference Proceeding publications (other than abstracts)

Donya Quick and Clayton T. Morrison. Composition by Conversation. In *Proceedings of the Forty-Third International Computer Music Conference* (ICMC), 2017. http://arxiv.org/abs/1709.02076

Colin R. Dawson, Chaofan Huang, and Clayton T. Morrison. An Infinite Hidden Markov Model with Similarity-biased Transitions. In *Proceedings of the Thirty-Fourth International Conference on Machine Learning* (ICML), 2017. https://arxiv.org/abs/1707.06756

(d) Manuscripts

(e) Books

(f) Honor and Awards

(g) Title of Patents Disclosed during the reporting period

(h) Patents Awarded during the reporting period

(2) Student/Supported Personnel Metrics **for this Reporting Period** (name, % supported, %Full Time Equivalent (FTE) support provided by this agreement, and total for each category):

(a) Number of Undergraduate STEM Students

(b) Number of Graduate STEM Students

Paul Hein, 100% supported, 100% FTE summer, MS in Computer Science, starting May 2017 (ongoing).

(c) Number of students that received a STEM degree

Paul Hein, 100% supported, 50% FTE academic year, BS in Computer Science, completed May 2017.

(d) Other Research staff (Name of each, FTE % Supported for each, Total % Supported) should be reported in the Participants section.

John Ivens, research programmer, 49% FTE.

Christopher Herald, PhD, music domain expert, 35% FTE.

(3) "Technology transfer" (any specific interactions or developments which would constitute technology transfer of the research results). Examples include patents, initiation of a start-up company based on research results, interactions with industry/Army R&D Laboratories or transfer of information which might impact the development of products.

None.

(4) Scientific Progress and Accomplishments (description should include significant theoretical or experimental advances)

A key development in the past year has been the addition of a second major use case for the MUSICA project: musical composition by conversation. In composition by conversation (hereafter, CBC), the MUSICA system interacts with human musicians through a natural language interface while human and MUSICA collaborate on the shared task of creating and editing a musical score. The ultimate goal is for MUSICA to be an equal participant in the composition process, suggesting creative additions and changes while understanding the intent of the human musician in the process of establishing shared goals. This use case is enabled by integrating the MUSICA system with the Cogent dialog agent framework (described below). Work this past year has also continued on the real-time music interaction use case aimed at providing MUSICA with the ability to interact with human musicians playing instruments, as in a trading-fours scenario (hereafter, T4).

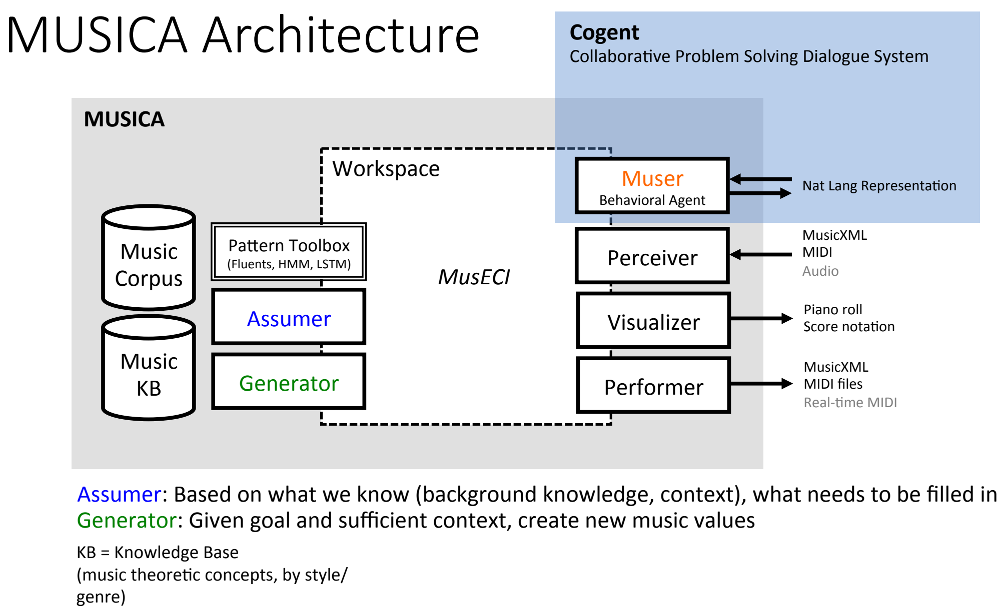

Figure 1: MUSICA Architecture

Figure 1 outlines the main components of the MUSICA architecture that have been developed over the past year. The architecture takes inspiration from a blackboard-style problem solving architecture in which a set of modular agents (called *knowledge sources*) are responsible for carrying out sub-tasks and collaborate by interacting with a *workspace* in which the overall task is incrementally completed. In the center of the Figure is MUSICA's Workspace, which is populated by representations of one or more musical scores and the states of problem solving in a CBC or T4 use case scenario. The objects in the Workspace are primarily represented in MusECI, a music knowledge and process representation language (described below). The boxes around the

Workspace represent MUSICA knowledge sources (which themselves comprise collections of algorithms for specific tasks). The Perceiver is responsible for reading music representations in several formats, translating them to MusECI representation. The Visualizer renders graphics of MusECI representations of musical scores. The Performer translates MusECI into music formats that can be played on instruments, for example through MIDI. The remaining four knowledge sources, Muser, Assumer, Generator, and the Pattern Toolbox, are described in more detail, below.

In order to accommodate the CBC and T4 use cases, MUSICA must represent music concepts in a form that supports conversation about musical objects as well as a way to represent patterns in music for recognition and generation. Beginning in fall 2016, the UA MUSICA team worked closely with Donya Quick of the SMU/Stevens team to develop MusECI, a language for representing **Mus**ical **E**lementary **C**omposable **I**deas. The language is designed to bridge several computational music paradigms, supporting music programming and algorithmic composition, querying and retrieving musical structure at varying levels of abstraction, and is influenced by cognitive models of musical representation to make it amenable to interfacing with natural language parsers. MusECI provides a representation of musical concepts that can support linguistic referential statements about musical objects and their relationships as well as performing operations to manipulate those objects and relationships. MusECI provides a set of symbolic primitives for representing musical objects such as notes and rests as compositions of pitch, beat, and onset properties, and a set of connecting primitive for composing these objects into parallel and sequential relationships in larger musical structures. The language incorporates a query interface for representing, searching and retrieving patterns of musical objects at different levels of abstraction, as well as a growing set of operations that provide methods for manipulating MusECI structures. The initial prototype design of the language is presented in Quick & Morrison (2017), and development and extension of the language is ongoing.

Figure 2 shows an example of a dialog that is currently supported by MusECI. The conversation starts with MUSICA ("Computer") assumed to have a MusECI representation of the score in the upper-right. The Human then utters sentences that are interpreted by MUSICA as MusECI statements, which when executed manipulate the score. The bottom of the Figure shows a schematic of the MusECI representation of the single-line melody as a sequence of notes, the pattern within the melody that the dialog references (dashed outline picking out particular notes), and the change to the MusECI structure as a result of the operations (changing F to G and C to B).

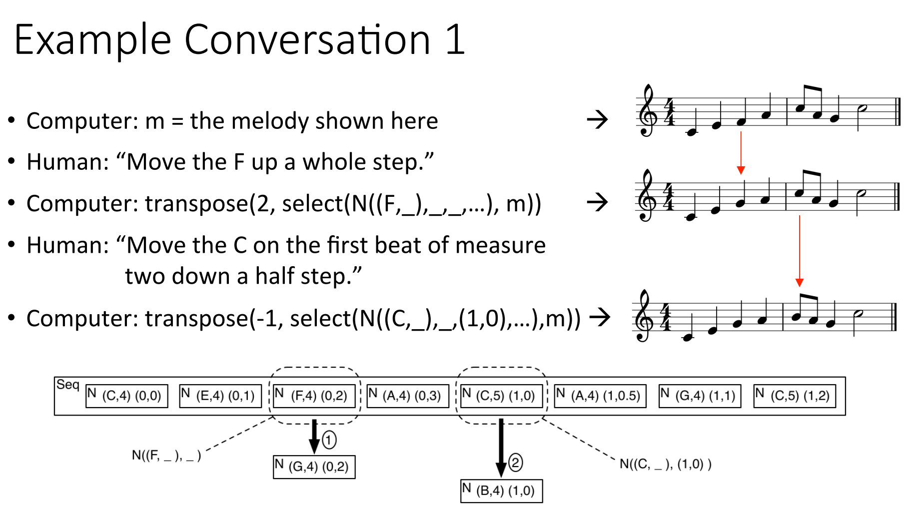

Figure 2: Example CBC showing score and manipulation of MusECI structure.

The dialog in Figure 2 demonstrates MusECI representation of linguistically specified musical structures, reference (selection patterns), and operations. But this particular exchange is constrained by the Human's utterances being unambiguously interpretable given the current score. Humans naturally rely on assumed background knowledge, implicature and pragmatics to communicate more efficiently while eliding details. The MUSICA team is developing the Assumer knowledge source as a reasoning agent that uses context and music background knowledge in an attempt to provide default values to resolve ambiguities. When the Assumer fails to confidently identify a default value, it will then engage the human conversant with targeted questions. Currently, the Assumer uses knowledge about musical key and scales to infer possible requested pitch changes when left unspecified, and also contains heuristics for reasoning about note lengths. The Assumer will continue to be a large focus of ongoing research.

In many situations, MUSICA will need to generate new musical objects. This will happen in the presence of variable amounts of context ranging from minimal-context requests such as, "give me two measures of music," to more constrained requests such as, "add a chord to the first beat of measure two." The Generator knowledge source is designed to handle the task of identifying the appropriate constraints to follow and then generate new musical structures within those constraints. This year the MUSICA team has adapted generation methods from the Kulitta generative grammar as well as explored machine learning approaches to learning generative models from analysis of transcribed music; the latter will be described below. Like the Assumer, the Generator will continue to be a key focus of ongoing research in the coming year.

In parallel to the development of MusECI, the UA MUSICA team worked with IHMC (particularly, Lucian Galescu and Choh Man Teng) to integrate MUSICA with the Cogent dialog agent framework (which includes the TRIPS parser). Integration involved defining music vocabulary corresponding to MusECI concepts and representing this vocabulary in the TRIPS ontology; this task is ongoing as the MUSICA team expands MusECI and the kinds of linguistic interactions MUSICA can accommodate. The MUSICA team created a dialog agent, Muser, which interfaces with Cogent and processes TRIPS logical form representations of sentence parses.

Figure 3 depicts the sequence of message exchanges among the Cogent dialog agents (boxes on the right-hand, blue side of Cogent) and Muser when processing a sentence input through the Cogent Keyboard.

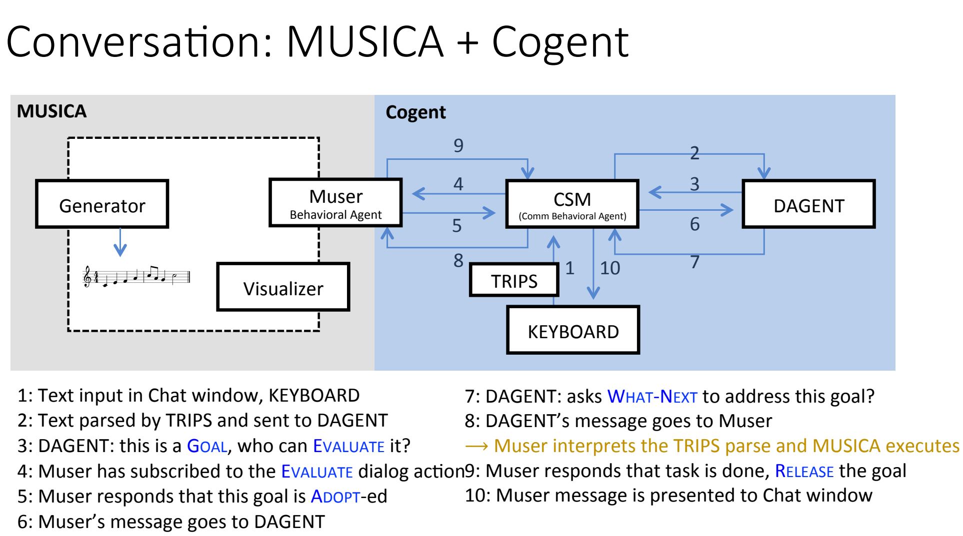

Figure 3: Muser and Cogent Dialog Interaction

Figure 4 outlines the steps involved in Muser interpreting the TRIPS logical form representation of a TRIPS parse of a natural language sentence; the steps in Figure 4 correspond to Step 8 in Figure 3. (1) represents the original sentence typed in by a human at the Keyboard. (2) represents the logical form graph produced by the TRIPS parse of the sentence. Portions of the tree are highlighted in color corresponding to the highlighted segments of the sentence in (1). Muser uses graph pattern-matching to identify these parse subgraphs, which in turn map to interpretations for MusECI representations, shown in step (3). Muser represents the utterance by first identifying the portion of the logical form that specifies the operation being *requested* (in this case, a transposition); next, Muser identifies the expression of the musical pattern in the MusECI score representation to select as the portion of the music that will be *affected* by the operation (selecting the first beat of measure two); finally, Muser identifies and maps additional information relevant to the specific operation parameters (in this case, the *result* of the operation is to move the pitch of the selected music down (-), and the *extent* to which it should be moved is 1 chromatic step). (4) Shows the resulting MusECI representation constructed from the information Muser extracts from the logical form graph. If the resulting MusECI representation does not contain enough information to execute the operation, there is then the opportunity to respond with targeted dialog with the user to clarify (this work is ongoing).

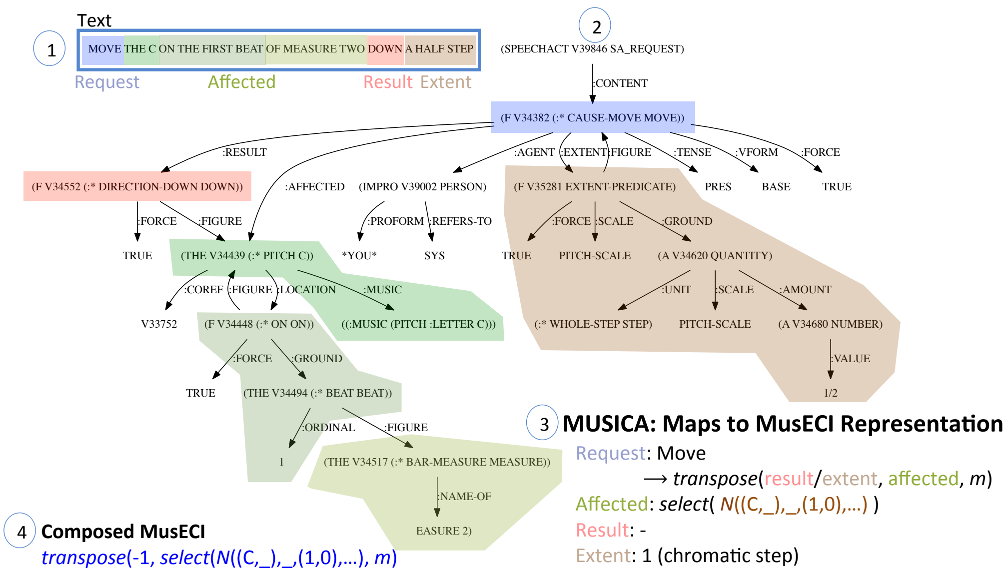

Figure 4: Muser mapping TRIPS logical form to MusECI

The final MUSICA knowledge source, the Pattern Toolbox, comprises a set of algorithms for analyzing music structures and building pattern models that can be used for generating music structures. In the following, we first review ongoing work on curating jazz scores that form the source data the Pattern Toolbox works on, then we review a new line of work on the Pattern Toolbox algorithms.

Over the past year, the UA MUSICA team has continued manual transcription of jazz scores into MusicXML by expert jazz musicians; the MusicXML representation is then translated automatically into MusECI. The MUSICA jazz corpus now comprises over 150 transcriptions, 6 of which are examples of two musicians engaging in trading-fours. The UA MUSICA team also worked with jazz expert musicians at UA and Stevens to carry out analysis of the trading fours scores. The goal of the analysis is to have the jazz experts use their musical training to identify musical patterns that they hear are being expressed by musicians, and identify when patterns appear to be picked up and adopted by a musician "responding" to what the other musician has played. Based on this analysis, we are developing a "vocabulary" of types of patterns and manually implementing these in the *fluent* framework (from year 1), so that MUSICA can then automate finding such fluents in new music scores. This work is ongoing.

While we strive to make the fluent patterns that MUSICA can recognize cover as many patterns as possible, their implementation is labor-intensive and may miss repeated patterns that a musician establishes while interacting in real-time in the T4 use case. We need an automated method for identifying repeated patterns. Such a method would not only potentially catch patterns in real time, but can also be used in searching over the jazz corpus for automating learning of new fluents.

Repeated patterns are the basis of *data compression*. When data contain a pattern that repeats, we can re-represent the data more compactly by noting the pattern once and then indicating where it reoccurs. Following this intuition, the UA MUSICA team has investigated data compression algorithms as a potential tool for identifying repeated patterns. The most promising algorithm the team has investigated is the SIATEC[1] algorithm for music score compression. The algorithm treats the score as a multi-dimensional dataset representing notes as points in beat (time) and pitch space. The overall method consists of two parts, with the SIA algorithm computing all maximal geometric patterns between notes, and SIATEC computing all occurrences of all the maximal repeated patterns in the dataset. The geometric patterns themselves consist of a set of relative changes in pitch and beat between a set of notes; if these relative relationships between notes can be found to repeatedly occur throughout the score, then the pattern is promoted. While this provides a basis for identifying pattern candidates, we have found that many candidates are not interesting. This is in part because the patterns that are identified are relatively unconstrained. The UA MUSICA team is extending the framework by incorporating additional constraints that serve to rank pattern candidates according to different criteria, including criteria that have been identified as part of the trading-fours score analysis. The team is evaluating these ranking methods on the MUSICA jazz corpus and also exploring ways to learn ranking criteria by rewarding criteria that find wide application in extracting patterns across the corpus. This work is ongoing.

---

[1] D. Meredith, K. Lemström and G. A. Wiggins. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, 31(4): 321-345.