A musical agent learns to generate a two-part invention using SARSAS. SARSAS is a reinforcement learning technique that learns an optimal policy by sampling the state space to estimate the utility of state-action pairs Q(s, a) where s denotes a state, a denotes an action, r denotes a reward, alpha denotes the learning rate and gamma the discount rate.

\[ Q(s, a) \leftarrow Q(s, a) + \alpha (r + \gamma Q(s', a) - Q(s, a)) \]

- First, the policy was learned using hand-crafted rules describing the desired characteristics of two-part inventions. These rules could also be discovered using data mining techniques.
- Then, the rules acted as a critic's comments to the generated music. The musical agent would amend its policy based on these comments.

In our approach, each episode was a complete 32-bar two-part counterpoint. Form and other contexts were incorporated into the system via the critic's rules and the usage of context dependent Q-tables.

In this work, we employed SARSAS to generate 32-bar two-part invention pieces. By carefully selecting the representation of states, actions, rules and contexts, a complex problem such as algorithmic composition could be dealt with and reasonable output could be obtained with comparatively less effort.

Policy learning in RL is a powerful concept. An agent explores a partially observable environment until it learns a policy (i.e., how it should react to the environment) that maximises its return, \$R\$. The representation of the state space, S, and actions, A, are critical since they are the abstraction of behaviours to be learned. In further work, the following directions could be pursued:

(i) to improve the handcrafted rules for different composition,
(ii) to automate rules-acquisition process, and
(iii) to apply the approach to other genres (e.g., four part writing, jazz, etc).

References:
- Sommik Pham-Amnellus: Generating Tonal Counterpoint Using Reinforcement Learning, ICDIN (1) 2009: 580-589