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

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

**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 maximizes its return, SRS. The representation of the state space, $S$, and actions, $A$, are critical since they are the abstraction of behaviors to be learned. In further work, the following directions could be pursued:

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

**In this work**, we employed SARSA 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.

**References**

- Somnuk Phon-Ainquanuk: Generating Tonal Counterpoint Using Reinforcement Learning. ICONIP (1) 2009: 580-589