We thank all reviewers for the constructive comments! We now respond to common and individual comments.

Common: Q1: On musical motivation and background. In the revision, we will give more introduction and references. 
A brief one: Counterpoint [1] is an essential and unique concept in Western music theory. Traditional Chinese music (e.g., folk songs[2] and operas), in its native form, does not have counterpoint practices, with few exceptions [3]. Some renowned Chinese composers, e.g., Xian Xinghai and He Luting, have explored incorporating counterpoints and fugues to Chinese music [4]. Notable works include “Yellow River Cantata” by Xian Xinghai and “Buffalo Boy’s Flute” by He Luting. However, systematic theories and broader influences on the general public of integrating counterpoint with Chinese folk melodies are lacking. This motivates our work. 

Q2: On the scope of application. Yes, the scope of the application could be widened in both (1) the music field and (2) beyond music in our future work. (1) Counterpoint patterns could be transferred to other non-Western music styles. The generation could contain more voices by evaluating the inter-rewarder between every two parts. (2) Our work could be applied to composition of other art forms, e.g., classical Chinese poems or couplets, which are rich of counterpoint-like patterns. 
Q3: On broader impact. In the revision, we will discuss more opportunities and challenges of cultural exchanges in both directions. For example, the inter-cultural style fused music could be used in Children’s enlightenment education to stimulate their interest in both cultures. We will also discuss other ways in which MIR may change music, following the suggested reference.

Q4: On causality. The generator’s observation strictly obeys the causality constraint. Line 152-153 and Figure 2 both show that: the generator will only observe the pitch and duration of the notes that end strictly before the onset of to-be-generated note, and only the pitch (but not duration) of the currently being played note. 
Q5: On the interaction reward. Yes, it refers to the “inter-rewarder”. The reason why FolkDuet has a higher interaction reward is due to the \( C(X_1, y_0^{(1)}) \) in Equation (3), which is different for the distribution of the notes in the initialization segments of FolkDuet and Bach chorales. Therefore, a high interaction reward does not necessarily mean a higher mutual information. 
Q6: On music notation in listening test. The music notation was shown to help subjects to attend the two voices. We used the notated key of the human part and the 4/4 time signature to render pitch spelling and bar line positions. There were some mistakes and subjects were informed to ignore such potential mistakes and focus on listening. 
Q7: On staff ordering in Figure 7. The model can generate both lower and higher voices for all music notation in the listening test and examples in the supplementary material, we put the machine part at the bottom regardless of their pitch height. 

Q8: On writing. Thanks! We will improve these in the revision. 

Q9: On the baseline. We did not compare with methods in [5, 14, 17, 24, 27], because (1) models in [17, 24, 27] are designed for monophonic melody generation instead of harmonization, and (2) rewards in [5, 14, 27] are rule-based and cannot appropriately capture the Chinese folk style or Bach chorale counterpoints.

---