

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

2 **Common: Q1: On musical motivation and background.** In the revision, we will give more introduction and references.  
3 A brief one: Counterpoint [1] is an essential and unique concept in Western music theory. Traditional Chinese music  
4 (*e.g.*, folk songs[2] and operas), in its native form, does not have counterpoint practices, with few exceptions [3]. Some  
5 renowned Chinese composers, *e.g.*, Xian Xinghai and He Luting, have explored incorporating counterpoints and fugues  
6 to Chinese music [4]. Notable works include “Yellow River Cantata” by Xian Xinghai and “Buffalo Boy’s Flute” by He  
7 Luting. However, systematic theories and broader influences on the general public of integrating counterpoint with  
8 Chinese folk melodies are lacking. This motivates our work. <sup>1</sup> **Q2: On the scope of application.** Yes, the scope of the  
9 application could be widened in both (1) the music field and (2) beyond music in our future work. (1) Counterpoint  
10 patterns could be transferred to other non-Western music styles. The generation could contain more voices by evaluating  
11 the inter-rewarder between every two parts. (2) Our work could be applied to composition of other art forms, *e.g.*,  
12 classical Chinese poems or couplets, which are rich of counterpoint-like patterns. **Q3: On broader impact.** In the  
13 revision, we will discuss more opportunities and challenges of cultural exchanges in both directions. For example, the  
14 inter-cultural style fused music could be used in Children’s enlightenment education to stimulate their interest in both  
15 cultures. We will also discuss other ways in which MIR may change music, following the suggested reference.

16 ■ **Reviewer 1: Q1: On the baseline.** Thank you for the suggestion on building another baseline with heuristics and  
17 rules. However, we find it not trivial to meet our online generation setting, and plan to attempt it in future work.  
18 **Q2: On whether the rewarder is fixed.** As the Chinese folk melodies for training are all monophonic, they are not  
19 appropriate to train or fine-tune the inter-rewarder; Adapting the inter-rewarder during RL training is likely to lose  
20 the counterpoint ground in the style-counterpoint balance. When polyphonic Chinese folk training data is available,  
21 we plan to adapt the inter-rewarder through few-shot learning. **Q3: On causality.** The generator’s observation strictly  
22 obeys the causality constraint. Line 152-153 and Figure 2 both show that: the generator will only observe the pitch and  
23 duration of the notes that end strictly before the onset of to-be-generated note, and only the pitch (but not duration) of  
24 the currently being played note. **Q4: On the 20 random pitch shifts.** The 20 pseudo duets for initializing the generator  
25 training are obtained by first randomly transposing the folk melody 20 times, and then creating a pseudo counterpart  
26 for each transposition through another random transposition. All transpositions cannot exceed the pitch range of the  
27 dataset. **Q5: On the interaction reward.** Yes, it refers to the “inter-rewarder”. The reason why FolkDuet has a higher  
28 interaction reward is due to the  $C(X_i, y_0^{(i)})$  in Equation (3), which is different for the distribution of the notes in the  
29 initialization segments of FolkDuet and Bach chorales. Therefore, a high interaction reward does not necessarily  
30 mean a higher mutual information. **Q6: On music notation in listening test.** The music notation was shown to help  
31 subjects to attend the two voices. We used the notated key of the human part and the 4/4 time signature to render  
32 pitch spelling and bar line positions. There were some mistakes and subjects were informed to ignore such potential  
33 mistakes and focus on listening. **Q7: On staff ordering in Figure 7.** The model can generate both lower and higher  
34 voices. For all music notation in the listening test and examples in the supplementary material, we put the machine  
35 part at the bottom regardless of their pitch height. ■ **Reviewer 2: Q1: On background questions in listening test.**  
36 Thank you for pointing this out and we admit that the background questions could be designed more carefully. In  
37 future work, we will survey subjects’ recency of music training. **Q2: On writing.** Thanks! We will improve these  
38 in the revision. ■ **Reviewer 3: Q1: On RL-Duet in listening test.** We considered adding RL-Duet in listening test.  
39 However, its Bach style is so strong that its generated parts differ sharply from the folk melodies. So, for the sake of  
40 resources, we excluded it in listening test. In the revision, we will make this clear in paper and show some examples  
41 in the supplementary and website. **Q2: On quality and reusability of Essen dataset.** Thank you for the suggestions!  
42 We admit that there are many performance-related features in Chinese folk music that cannot be recorded in the MIDI  
43 transcriptions. That being said, the typical transcription format of Chinese folk music is the numbered musical notation  
44 (plus occasional playing technique marks such as glissandi), which is almost equivalent to MIDI. This suggests that  
45 these performance-feature-removed transcriptions still convey the Chinese folk style. Therefore, we believe that our  
46 symbolic generation work is well grounded on this dataset. In the revision, we will provide a clearer description and  
47 justification. For the Essen dataset, the “Han, Natmin, Shanxi, and Xinhua” classification comes from the original  
48 dataset and also confuses us. In the revision, we will add a footnote and refer readers to the original dataset for more  
49 information :) **Q3: On beat synchronization.** We calculate beat locations of notes with the 4/4 time signature assumption  
50 and 16th note quantization. Take notes in Figure 2(a) as an example, their onsets are, in 16th note indexes, at 0, 2, 4, 6,  
51 8, 16, 20, 24, 26, 28 and 30. Their beat locations are then at 0, 2, 4, 6, 8, 0, 4, 8, 10, 12 and 14. As many pieces in the  
52 Essen dataset use other time signatures and may even vary them within a piece, we admit that our simple treatment may  
53 confound the model training, and we plan to improve this in future work. **Q4: On writing.** Thanks! We will fix them all  
54 in the revision. ■ **Reviewer 4: On the baseline.** We did not compare with methods in [5, 14, 17, 24, 27], because (1)  
55 models in [17, 24, 27] are designed for monophonic melody generation instead of harmonization, and (2) rewards in [5,  
56 14, 27] are rule-based and cannot appropriately capture the Chinese folk style or Bach chorale counterpoints.

<sup>1</sup>[1] Laitz. The complete musician. [2] Jones, et al. Folk music of China: Living instrumental traditions. [3] Wiant. Possibilities of Polyphonic Treatment of Chinese Tunes. [4] Ouyang, et al. Ideology and National Identity in 20th-Century Chinese Music.