A comparison of melody interpolation performed by human and artificial intelligence based on human similarity judgments

Hyunjae Kim (a), Eunji Oh (a), Jieun Park (a), Yoonjin Chung (b), Juhan Nam (a, b), Kyung Myun Lee* (a, c)

(a) Graduate School of Culture Technology, KAIST, Republic of Korea, *kmlee2@kaist.ac.kr
(b) Kim Jaechul Graduate School of Artificial Intelligence, KAIST, Republic of Korea
(c) School of Digital Humanities and Computational Social Sciences, KAIST, Republic of Korea

Keywords: Melody similarity, melody interpolation, artificial intelligence, automatic composition

Introduction

Automatic composition models based on artificial intelligence (AI) techniques have made it possible to automatically generate musical melodies in the style of a particular artist. Recently, models have also created new melodies by mixing and interpolating two different melodies into intermediate levels [1]. Given the issue of copyright, the relationship between interpolated melodies and original reference melodies is important, but the perceptual similarity between them has rarely been examined [2]. In this study, we aimed to investigate the perceptual similarity between interpolated and reference melodies in terms of contour and tonality, two important components of melody perception [3], by conducting a behavioral experiment. We also aimed to compare the composition mechanism of AI and human composers by analyzing interpolated melodies created by each.

Methods

A professional human composer and an AI composer created three sets of interpolated melodies consisting of eight isochronous notes from two reference melodies, which were at the extreme ends of contour and tonality. The interpolated melodies were composed at three intermediate levels of the two reference melodies with the intention of making gradual changes at each level of contour or tonality. For the AI composition, we used an explicitly-constrained conditional variational autoencoder (EC2-VAE), which interpolated the latent feature space of the pre-trained model trained with British and American folk songs. In the experiment, the interpolated melodies were randomly presented with the two reference melodies, and 33 participants rated the similarity between the intermediate and reference melodies on a scale of 0–100. Linear regression was then used to analyze the relation between the level of interpolation and the melodic similarity rated by the participants.

Results

All regression models for the AI and human composers are statistically significant. For contour interpolation, the human composer model ($\mathbf{R^2 = 0.858},\ p < .001, F = 2618.$) more reliably explains the gradual increase in similarity score along the interpolation steps than the AI composer model ($\mathbf{R^2 = 0.667},\ p < .001, F = 865.8$). For tonality, conversely, the AI composer model ($\mathbf{R^2 = 0.781},\ p < .001, F = 1548.$) shows a higher R-square score than the human composer model ($\mathbf{R^2 = 0.704}, p < .001, F = 1028.$). Additionally, different criteria metrics, such as mean-square-error (MSE) and mean-absolute-error (MAE), showed consistent results for the AI and human composers: MSE = (200., 491.) and MAE = (11., 16.) for contour, and MSE = (16., 12.) and MAE = (428., 284.) for tonality, respectively.

Conclusion

The results of the human similarity judgments showed that both interpolations generally reflected intermediate levels of the two reference melodies well. In particular, the AI interpolation better showed gradual changes in tonality between the two references, whereas the human interpolation better reflected gradual contour steps. This result suggests that AI and human composers have different interpolation styles. However, given the limited scope of our experiment with a single composer and AI model, further investigation is needed to include a wider range of human composers and different AI composition models with a diverse set of reference melodies.

References

Yang, R., Wang, D., Wang, Z., Chen, T., Jiang, J., & Xia, G. (2019). Deep music analogy via latent representation disentanglement. arXiv preprint arXiv:1906.03626.

Vempala, N. N., & Russo, F. A. (2012, July). A melodic similarity measure based on human similarity judgments. In Proc. of Int. Conf. on Music Perception and Cognition (ICMPC).

Schmuckler, M. A. (2016). Tonality and contour in melodic processing.