EXPLORING SAMPLING STRATEGIES IN LATENT SPACES FOR MUSIC GENERATION

Nádia CARVALHO (nscarvalho@fe.up.pt) (0000-0001-6882-5138)[1] and
Gilberto BERNARDES (gba@fe.up.pt) (0000-0003-3884-2687)[1]

[1]**University of Porto – Faculty of Engineering and INESC TEC**, Porto, Portugal

ABSTRACT

This paper investigates sampling strategies within latent spaces for music generation, focusing on (chordified) J.S. Bach Chorales and utilizing MusicVAE as the generative model. We conduct an experiment comparing three sampling and interpolation strategies within the latent space to generate chord progressions – from a discrete vocabulary of Bach's chords – to Bach's original chord sequences. Given a three-chord sequence from an original Bach chorale, we assess sampling strategies for replacing the middle chord. In detail, we adopt the following sampling strategies: (1) traditional linear interpolation, (2) k-nearest neighbors, and (3) k-nearest neighbors combined with angular alignment. The study evaluates their alignment with music theory principles of functional harmony embedding and voice-leading to mirror Bach's original chord sequences. Preliminary findings suggest that k-nearest neighbors and k-nearest neighbors combined with angular alignment closely align with the tonal function of the original chord, with k-nearest neighbors excelling in bass line interpolation and the combined strategy potentially enhancing voice-leading in upper voices. Linear interpolation maintains aspects of voice-leading but confines selections within defined tonal spaces, reflecting the non-linear characteristics of the original sequences. Our study contributes to the dynamics of latent space sampling for music generation, offering potential avenues for enhancing explainable creative strategies.

1. INTRODUCTION

The rise of generative models in music composition has sparked a notable interest in delving into the latent spaces of musical data for creative pursuits [1–3]. As AI-driven technologies advance, there is a growing curiosity about the intricate mechanisms underlying these models and their potential to unlock new realms of musical expression. Latent spaces, in particular, present an enticing prospect for composers and researchers, offering a rich landscape where musical ideas can be explored, manipulated, and synthesized in innovative ways [4–6]. This fascination has led to investigations into the dynamics of latent space navigation, aiming to unveil the hidden structures governing musical composition and interpretation [5, 7–9].

The pursuit of uncovering meaningful music semantics within latent spaces for controllable sampling based on distinct music characteristics is a vibrant area of investigation within the realms of music informatics and artificial intelligence [9–12]. This research endeavor is driven by the overarching goal of equipping users, musicians, and composers with sophisticated tools, enabling them to engage in interactive exploration and creation of music tailored to their preferences and specifications. The potential applications of such advancements span a broad spectrum, encompassing domains like music composition, production, interactive entertainment, and personalized music recommendation systems [9].

Exploring a latent space involves navigating its multi-dimensional structure to capture the essence of music semantics. This capability empowers users to transition between different representations, each embodying a unique combination of musical traits, facilitating the exploration of a multitude of musical variations and possibilities. Additionally, interpolation, a traditional method of exploring latent spaces, involves seamlessly transitioning between established music representations within the latent space, thereby facilitating smooth transitions and novel combinations [13, 14].

In this context, our study delves into the realm of music generation, seeking to elucidate the role of sampling strategies within latent spaces, with a particular focus on J.S. Bach Chorales as a benchmark test set. Our primary objective is to evaluate the efficacy of different sampling and interpolation strategies within the latent space, using MusicVAE [15] as our generative model.

The fundamental aim of our evaluation is to ascertain the extent to which these strategies can accurately generate chord progressions that adhere to Bach's original sequences. To achieve this, we conduct an experiment comparing three distinct sampling strategies. These strategies involve departing from a chorale's chord progression and exploring chord substitutions based on the previous and sequential chord contexts. For instance, given a chord sequence [A, B, C], our goal is to generate new sequences such as [A, B', C], [A, B'', C], and so forth.

The three sampling strategies under scrutiny include the traditional linear interpolation between adjacent chords, k-nearest neighbors, and a novel approach combining k-nearest neighbors with angular alignment to $\vec{AB}$. By adopting these strategies, we aim to assess their efficacy

in preserving the essence of Bach's style while introducing variations that retain semantic coherence within the latent space.

Through our investigation, we aim to provide insights into manipulating latent spaces for music generation, enhancing creative autonomy and stylistic fidelity in algorithmic composition by examining different sampling strategies. Our methodology integrates music theory to assess the similarity between generated sequences and original Bach chorales, allowing us to evaluate stylistic fidelity and delve into latent space semantics. Ultimately, our study seeks to contribute to the ongoing discourse surrounding the intersection of AI and music, paving the way for more sophisticated and nuanced approaches to automated composition.

Our paper is structured as follows. In Section 2, we summarize the sampling strategies employed, namely: (1) traditional linear interpolation between adjacent chords, (2) k-nearest neighbors, and (3) k-nearest neighbors combined with angular alignment to $\vec{AB}$. Section 3 outlines the methodology employed to assess the effectiveness of the sampling strategies in generating chord progressions adhering to a specific musical style. This section provides comprehensive details regarding the experimental setup, including the dataset and the model utilized for the evaluation. Subsequently, Section 4 presents and discusses the obtained results. Finally, Section 5 summarizes the conclusions drawn from our study and suggests potential avenues for future research.

2. SAMPLING STRATEGIES

In this section, we examine sampling techniques enabling us to traverse a discrete set of chords depicted in the latent space and produce novel sequences from a specified chord progression [A, B, C], yielding variations like [A, B', C], [A, B", C], and so on. The pursued strategies include: (1) conventional linear interpolation between consecutive chords, (2) k-nearest neighbors, and (3) a novel approach integrating k-nearest neighbors with angular alignment to $\vec{AB}$.

The subsequent sections delve into each of these strategies in intricate detail, elucidating their principles, methodologies, and implications in the context of our exploration of latent space for music generation.

2.1 Linear Interpolation

Conventional generative music methodologies using VAE latent spaces [13, 14, 16] often entail sampling via linearly interpolated coordinates between two designated musical elements. Musical elements, such as chords, measures, or phrases, can adopt different time scales. Typically, a predetermined number of points are sampled at equal intervals between the starting and ending points, such that:

$$\vec{p} = (1-t)\cdot\vec{p}_0 + t\cdot\vec{p}_1 \quad , \tag{1}$$

where $\vec{p}$ is the interpolated point, $\vec{p}_0$ and $\vec{p}_1$ are the two points between which interpolation is being performed, and t is the interpolation parameter ranging from 0 to 1, determining the weight of $\vec{p}_1$.

This methodology aims to facilitate smooth transitions between predefined musical sequences, predicted on the assumption of a continuous musical spectrum [17]. We sample the dataset to pinpoint the chord nearest to the equally spaced interpolated coordinate to align with our dictionary-constrained chord space.

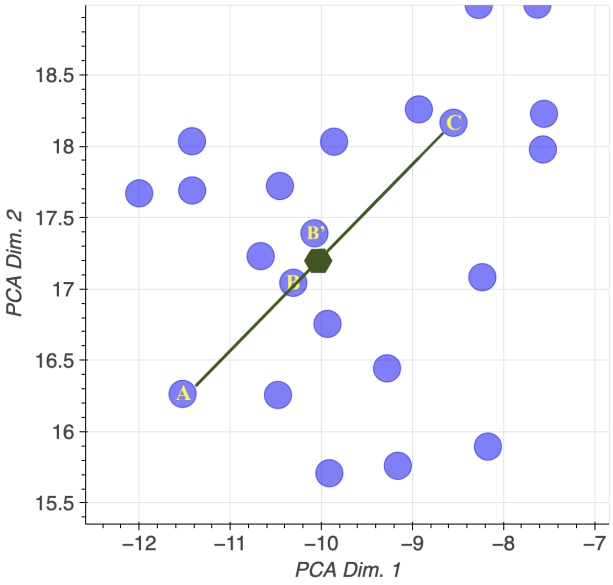


Figure 1. Sampling a chord B' from the latent space (represented in a two-dimensional space) using traditional Linear Interpolation from a sequence of selected chords [A, B, C]. Blue circles represent discrete chords projected in the latent space. The midpoint, found via linear interpolation between A and C, is shown as a dark-green hexagon.

Figure 1 depicts the interpolated sampling technique. Starting from points A and C within the latent space, we interpolate a single point by interpolating linearly between A and C, with a t halfway between 0 and 1. The blue circles in the illustration denote discrete chords projected in the latent space. The midpoint, determined through linear interpolation between A and C, is represented by a dark-green hexagon. In this case, the nearest discrete point is, coincidentally, the original chord B. Thus, the sampled chord, labeled B' in Fig. 1, corresponds to the second nearest discrete chord point.

2.2 k-Nearest Neighbors

The k-nearest neighbors (k-NN) algorithm is a versatile approach commonly employed in classification and regression tasks within supervised learning contexts [18]. This algorithm functions by determining the value of a data point based on the average value of its k nearest neighbors in the feature space. This intuitive principle allows k-NN to adapt effectively to diverse datasets [19].

In this approach, we rely on the perceived significance of the latent space to identify chords that are similar to a given target chord based on their proximity in the embedding space. To sample a chord from the corpus chords

represented in the latent space, we utilize k-NN to choose the k nearest neighbors of a chord within a predefined embedding space. This process, illustrated in Figure 2, involves computing distances (e.g., Cosine or Euclidean) between the target chord and all other chords in the embedding space and selecting the k nearest neighbors. These nearest neighbors serve as candidates for sampling, providing a diverse set of potential chords that are expected to be related to the target chord.

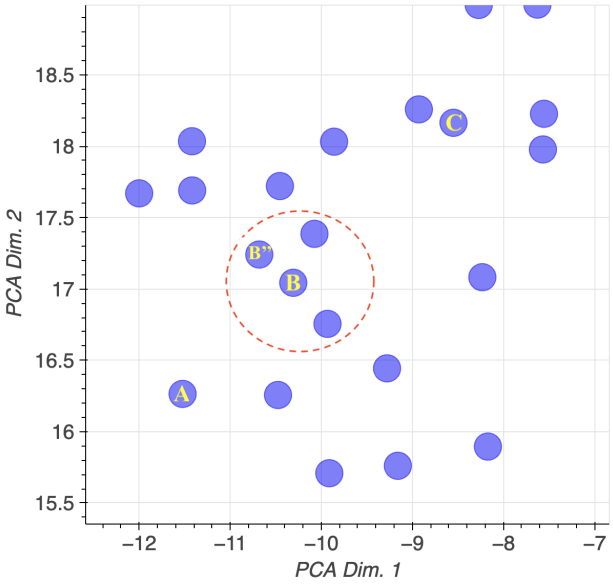


Figure 2. Sampling a chord B" from the latent space (represented in a two-dimensional space) using k-NN. Blue circles represent discrete chords projected in the latent space. The k nearest neighbors (in this case, $k = 3$) are identified within the dotted red line.

In this case, as displayed in Figure 2, we only take B into consideration to find its substitute chord. We search within the nearest neighbors of B (circles within the dotted red line in the Figure; in this case, we use $k = 3$). The closest candidate, according to a measure of distance (e.g., Euclidean or cosine), is notated as B".

2.3 k-Nearest Neighbors and Angular Displacement

We depart from our previous strategy by suggesting a fusion of k-NN with angular alignment, meticulously crafted for the chord pair vector $\vec{AB}$. Angular displacement, used frequently to improve the handling of high-dimensional data whose direction is important [20], involves aligning vectors in a multi-dimensional space according to their angles. Through employing angular alignment to the original chord sequence, our aim is to preserve the overarching direction of the initial sequence.

The initial phase of this sampling strategy mirrors the previous approach: we examine the nearest neighbors of B and curate a subset of the k closest ones using a designated distance metric such as Cosine or Euclidean distance. In this instance, k must encompass a broader selection of chords since we exclusively apply angular alignment to this subset. To compute the angular displacement relative to the vector $\vec{AB}$, we first determine the vector itself. Next, we compute the vector from A to each potential chord point within the set of the k closest ones. Subsequently, we assess the angular disparity between each of these vectors and the original one. The chord displaying the least angular deviation is identified as the optimal replacement.

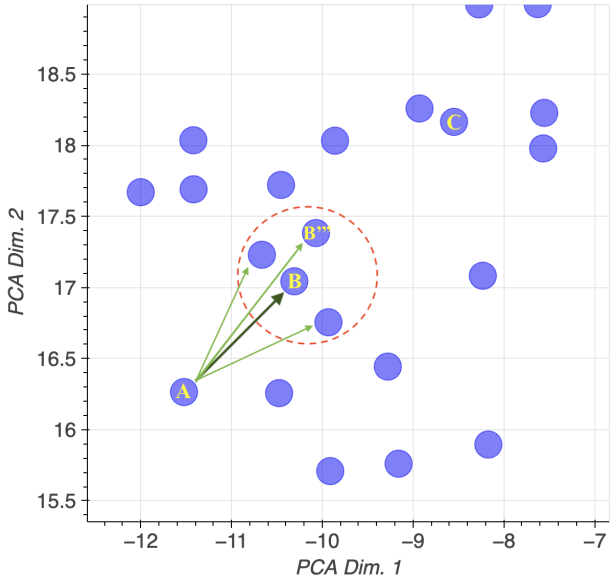


Figure 3. Sampling a chord B"' from the latent space (represented in a two-dimensional space) using k-NN with Angular Alignment to the vector formed from the latent space representation of the selected sequence of chords [A, B]. In this visualization, discrete chords are depicted as blue circles. Within the dotted red line, the k nearest neighbors (in this instance, $k = 3$) are identified. A dark-green arrow indicates the vector from chord A to B, while light-green arrows represent the vectors from chord A to each candidate point.

In Figure 3, we showcase an example wherein we identify the trio of closest neighbors of B and calculate the angular separation between vector $\vec{AB}$ and the extension from A to each prospective point. Even though it isn't the closest neighbor, the closest candidate chord, ensuring optimal continuity, would be the one marked as B"' in the illustration due to its minimal angle.

3. EVALUATION

In this section, we thoroughly evaluate different sampling strategies in latent spaces for generating music, focusing specifically on (chordified) J.S. Bach Chorales as our main test set. Our main goal is to carefully examine various sampling and interpolation techniques within the latent space, using MusicVAE as our model.

Our evaluation seeks to gauge the effectiveness of various sampling methods in generating chord progressions akin to Bach's original compositions. We compare three sampling strategies as outlined in Section 2: one based on the conventional linear interpolation method, drawn from existing literature, and two innovative strategies centered around

the concept of perceptual relatedness shared by neighboring points in the latent space.

Our hypothesis posits that sampling within a latent space should capture key musical principles such as chord substitutions and parsimonious voice leading. We anticipate a cumulative effect from our strategies: interpolating between adjacent chords introduces variations that harmonize seamlessly with the original musical style. Additionally, employing k-NN enables us to identify chords that share similar traits with the original chord, maintaining harmonic coherence. This approach facilitates the exploration of potential chord substitutions while preserving continuity and smooth voice leading within the progression. This technique is expected to further ensure the preservation of musical continuity and smooth voice leading when coupled with angular displacement to a chord sequence. Moreover, angular displacement is expected to enhance the perceptual similarity to the original chord sequence.

3.1 Materials

To construct a latent space using MusicVAE, we utilize an architecture incorporating a recurrent encoder featuring a two-layer bidirectional LSTM with 1024 units. The decoder is structured hierarchically, employing a two-layer unidirectional LSTM with a hidden size of 1024 for both the conductor and decoder components. A latent size of 256 is selected to significantly reduce input dimensionality while preserving adequate information for effective reconstruction [14]. During the training phase, we utilize the Adam optimizer with an initial learning rate set to 10E-4 and a batch size of 16 to minimize the loss function and refine the model parameters.

As input to our model, we utilize a binary 128-element vector piano roll encoding, representing note activations per unit of time as 'salami slices,' [1] to capture intricate pitch changes within the music.

We employ a standard benchmark of 371 chorales by J.S. Bach to evaluate the proposed strategies, which act as a representative tonal music corpus. The stylistic nature of Bach's chorales makes them an ideal corpus to study harmonic information within the VAE latent space, as the remaining textural aspects are somehow constant, e.g., number of voices, instrumentation, limited change in pitch registers, and harmonic rhythm.

These chorales are obtained from the `music21` library. [2] The corpus comprises chorales in major (53%) and minor (47%) keys, with G Major (14%), A minor (12%), and G minor (11%) being the most prevalent tonalities. On average, each chorale consists of 84 chordified slices. Our training process utilizes 60% of the chorales from the corpus, while the remaining 40% are reserved for testing. To enhance diversity, the chorales utilized for training are subject to augmentation through transposition across all twelve keys, ascending (transposing to the upper 6 keys), as well as descending (transposing to the lower 6 keys).

[1] These slices entail segmented information, capturing the addition or subtraction of pitches from the musical surface each time a change occurs. The salami slice segmentation can be computed from MIDI files by the `chordify()` function within the `music21` software package.

[2] `http://web.mit.edu/music21/`, accessed on March 10, 2024.

3.2 Method

Utilizing the trained VAE latent space, we derive regional chord sequences directly from the original dataset, without transposed augmentation. We gauge the proximity of the three sampling strategies delineated in Section 2, operating under the assumption that the closer they align with the original dataset, the higher their efficacy.

In total, we gather 15276 three-chord sequences from major chorales, with the possible dataset comprising solely chords within major chorales. Furthermore, 13324 three-chord sequences are sourced from minor chorales, exclusively considering chords existing within these chorales for the discrete dataset. Lastly, 28600 three-chord sequences are extracted from all chorales, encompassing all chords within the chorales for the discrete dataset.

We utilize chord distances within a perceptually inspired pitch-space to determine the sampling strategy closest to the original [21]. Here, the Euclidean distances between pitch-class sets reflect their perceptual relatedness [22]. Chords are projected as a chroma vector c_n into a weighted pitch-class Discrete Fourier Transform (DFT) space, using Equation 2, where magnitudes and phases correspond to chord qualities and regional sets. N = 12 is the dimension of the chroma vector, and w_q are weights derived from empirical dissonance ratings of dyads used to adjust the contribution of each dimension q of the space [23].

$$T(q) = w_q \sum_{n=0}^{N-1} \bar{c}_n e^{\frac{-j2\pi kn}{N}}, \quad 1 < q < 6 \qquad \text{with} \quad \bar{c}_n = \frac{c_n}{\sum_{n=0}^{N-1} c_n} \tag{2}$$

In this space, we can compute the Euclidean distance between two given $T_1(k)$ and $T_2(k)$ vectors, representing chords, by utilizing Equation 3 [21].

$$d\{T_1(k), T_2(k)\} = \sqrt{\|T_1(k) - T_2(k)\|} = \sqrt{\sum_{k=1}^{M} |T_1(k) - T_1(k)|^2} \tag{3}$$

We analyze descriptive statistics for the 15276 major, 13324 minor, and 28600 combined chords. The sampling strategy with the lowest mean value serves as an indicator of the closest resemblance to the baseline original J.S. Bach chord sequences.

4. RESULTS AND DISCUSSION

The results of our evaluation are shown in Table 1. Regardless of the distance measures employed and the chorale sets evaluated, chords sampled via the k-NN method consistently exhibit the closest proximity to the original. Following closely are the chords sampled using the k-NN with

angular displacement technique. However, despite our expectations, the angular displacement fails to contribute significant information towards capturing chords that exhibit greater perceptual similarity to the original chord. Linear interpolation consistently yields the poorest results, often exhibiting distance values nearly twice as high compared to the other sampling methods.

		Major	Minor	All
Euclidean Distance	Linear Interp.	19.6 ± 6.8	20.5 ± 6.8	20.1 ± 6.8
	k-NN	**9.6** ± 4.3	**10.2** ± 4.5	**9.5** ± 4.3
	k-NN w/ AD	10.0 ± 4.6	10.7 ± 4.8	9.9 ± 4.5
Cosine Distance	Linear Interp.	19.7 ± 7.2	20.6 ± 7.2	20.1 ± 7.2
	k-NN	**9.5** ± 4.2	**10.2** ± 4.4	**9.4** ± 4.2
	k-NN w/ AD	10.0 ± 4.6	10.8 ± 4.8	9.8 ± 4.5

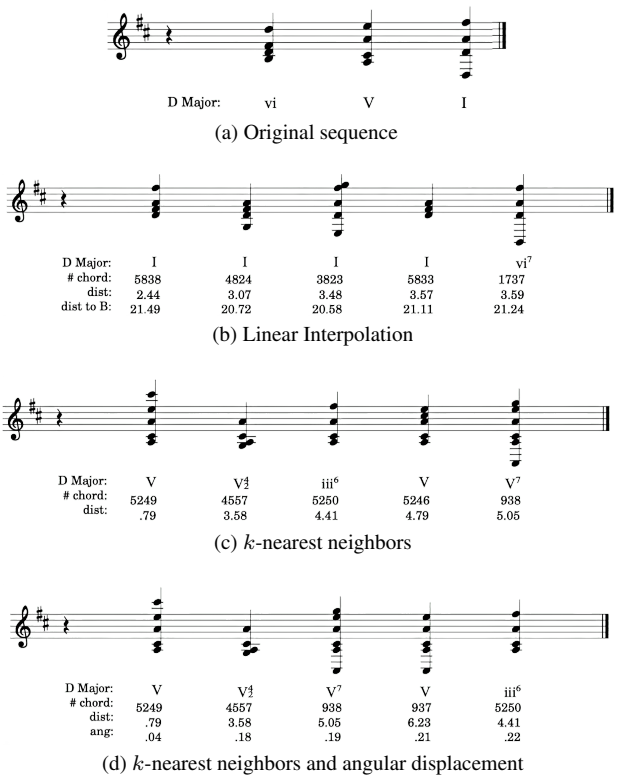
Table 1. Results of the comparative analysis regarding the perceptual relatedness of the chord sampling strategies. Descriptive statistics for 15276 major, 13324 minor, and 28600 combined chords are presented. Three strategies — Linear Interpolation, k-NN, and k-NN with angular displacement — are evaluated. Distances within strategies, calculated using both Euclidean and cosine measures, are depicted. Best results per dataset of chords are highlighted in bold.

To further inspect the quality of these sampling strategies and enlighten the raised hypothesis that a latent space ought to capture key principles from music theory and practice, such as chord substitutions (by replacing a given chord in a sequence and maintaining the same function) and similar voice leading between adjacent chords in a given sequence, we will study in detail two three-chord sequences from BWV 184.5 and BWV 311, chorales in major and minor keys, respectively.

Figure 4 illustrates a three-chord sequence corresponding to the (authentic) cadence of the initial phrase in J.S. Bach's Choral BWV 184.5 in D Major, along with the five closest candidates selected by each sampling strategy, ordered by closeness to the original. Notably, both the k-NN and k-NN with angular displacement strategies demonstrate comparable selections for four of the five candidate chords, indicating similar hypotheses underlying their choices.

Intriguingly, the third and fifth options are interchanged between the two strategies, while the fourth option varies from k-NN to k-NN with angular displacement. Nevertheless, all options remain within the dominant space, maintaining the function of the original chord.

In terms of voice-leading within the original sequence,



(a) Original sequence

(b) Linear Interpolation

(c) k-nearest neighbors

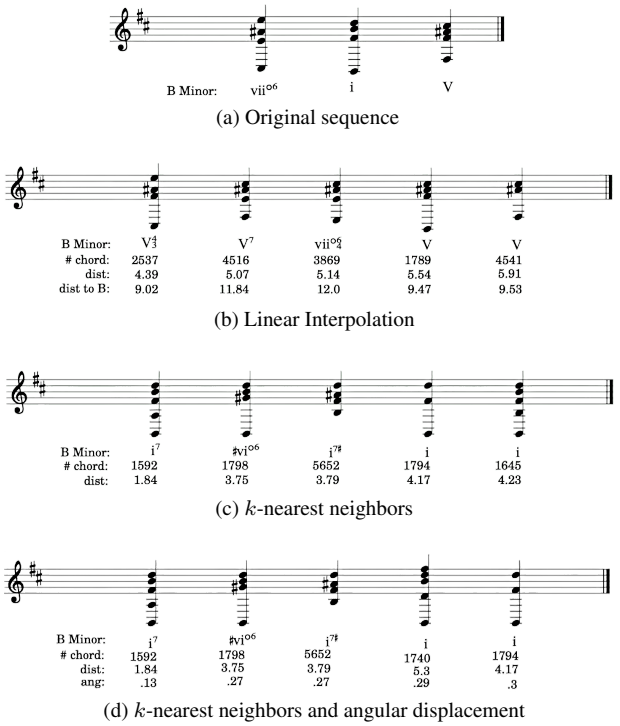
(d) k-nearest neighbors and angular displacement

Figure 4. Original sequence, derived from the cadence of the first phrase of J.S. Bach's Choral BWV. 184.5 in D Major, along with 5 optimal sampling candidates using the three different sampling strategies, ordered by closeness to the original. For each strategy, we present the tonal functions within the phrase, the number of the chord in the dataset of all chords, and the distance metric calculated, i.e., in k-NN and k-NN with angular displacement, we present the Euclidean distance (dist) from the candidate point to the original, plus angular distance for k-NN with angular displacement (ang). For linear interpolation, we present the Euclidean distance from the candidate point to the interpolated continue point in the latent space (dist) and the distance to the original chord (dist to B).

contrary to our anticipation, k-NN appears to yield superior bass line interpolation, as the bass notes of the initially sampled chords are notably closer to the first chord. Conversely, diverse candidate chords in k-NN with angular displacement could potentially enhance voice-leading in the upper voices. Even if the bass line leaps from the first chord in the progression to one of these chords, it remains stylistically acceptable, given that Bach's chorales often feature bass line jumps. Additionally, the substantial difference in bass between the first and third chords in the progression suggests that the bass leap from the second chord could effectively prepare the transition to the third chord, aligning with typical movement in Bach's chorales. For instance, both the third and fourth potential selections of k-NN with angular displacement present intriguing options for voice-leading, with the lower and upper voices resolving in opposing directions towards the final chord of the sequence.

In the case of linear interpolation, a contrasting pattern

emerges: all candidates fall within the tonal space of the tonic, notably closer to both the first and third chords than to the second chord in the original sequence. We posit that this phenomenon occurs because the original sequence does not adhere to a linear function, where the middle chord is functionally more distant from both the preceding and succeeding chords. Thus, the interpolated point will tend to remain closer to the starting and ending points of the linear function. Consequently, the movement between voices in all chords is much closer than in the last two strategies, as the notes are mainly the same between the three chords.



(a) Original sequence

(b) Linear Interpolation

(c) k-nearest neighbors

(d) k-nearest neighbors and angular displacement

Figure 5. Original sequence, derived from the cadence of the first phrase of J.S. Bach's Choral BWV 311 in B Minor, along with 5 optimal sampling candidates using the three different sampling strategies, ordered by closeness to the original. For each strategy, we present the tonal functions within the phrase, the number of the chord in the dataset of all chords, and the distance metric calculated, i.e., in k-NN and k-NN with angular displacement, we present the Euclidean distance (dist) from the candidate point to the original, plus angular distance for k-NN with angular displacement (ang). For linear interpolation, we present both the Euclidean distance from the candidate point to the interpolated continue point in the latent space (dist) and the distance to the original chord (dist to B).

In the minor domain, we introduce a second example, displayed in Figure 5. This figure showcases a chord progression referencing a half cadence, ending the initial phrase of J.S. Bach's Choral BWV. 311 in B Minor.

In this scenario, a familiar pattern emerges: the linear interpolation strategy yields candidate chords with voice leading closer to the first and last chords of the original sequence, yet the preservation of tonal function is compromised (in this instance, remaining within the dominant domain). Nonetheless, the difference between these candidates and those in the previous example is less pronounced, with the Euclidean distance to the original chord being notably closer, albeit still farther from the alternatives presented by the other two strategies.

The k-NN and k-NN with angular displacement strategies exhibit comparable outcomes to those in the previous example, with similar options across the two methods, barring one exception. However, between the two different chords proposed by the strategies, the one selected by k-nearest neighbors would likely provide superior voice leading in this specific sequence. This preference arises as the voice leading of the chord suggested by angular displacement would introduce a fourth interval in the upper voice between the second and third chords of the progression, a departure from the natural style observed in Bach's chorales.

5. CONCLUSIONS AND FUTURE WORK

This paper delves into the exploration of sampling strategies within latent spaces for music generation, specifically focusing on chordified J.S. Bach Chorales as a benchmark test set and utilizing MusicVAE as the generative model. We experimentally compare three sampling and interpolation strategies within the latent space to generate chord progressions mirroring Bach's original chord sequences. These strategies depart from an original Bach chorale phrase and evaluate chord substitutions given the previous and sequential chord contexts. Specifically, we employ the following sampling strategies: (1) traditional linear interpolation between adjacent chords, (2) k-nearest neighbors, and (3) k-nearest neighbors combined with angular alignment.

Our approach investigates these sampling strategies' alignment to music theory principles of functional harmony embedding and voice-leading to assess the similarity to the original Bach style and the semantic description of the space. The study yields valuable insights into the dynamics of latent space manipulation for music generation, paving the way for potential advancements in creative autonomy and stylistic fidelity in algorithmic music composition.

Our preliminary findings suggest that k-NN and k-NN with angular alignment to $\vec{AB}$ most closely align with the tonal function of the original chord, especially the former. Interestingly, k-NN tends to offer superior bass line interpolation, while k-NN with angular displacement potentially enhances voice-leading in upper voices. Despite slight variations between strategies, they generally maintain the essence of the original chord progression while offering alternatives with varying degrees of voice-leading quality and tonal function preservation.

Conversely, linear interpolation seems to excel in maintaining aspects of voice-leading. This could stem from its incorporation of both preceding and succeeding chords as reference points, resulting in smoother transitions and minimized leaps in the voices. However, linear interpolation often confines its selections within the tonal space defined

by these two chords. We hypothesize that this tendency may arise from the non-linear characteristics inherent in the original sequence.

These findings underscore the complexity of generating plausible chord progressions within the context of Bach's chorales and highlight the nuanced differences between sampling strategies regarding voice-leading and tonal function preservation. The study suggests that while each strategy offers viable alternatives, considerations such as bass line interpolation, upper voice resolution, and adherence to Bach's stylistic conventions play crucial roles in determining the most suitable chord progression.

In forthcoming research, motivated by the revelation that linear interpolation does not fully capture the original chord progressions in Bach's chorales, we plan to explore alternative interpolation strategies rooted in the broader framework of parametric curves. Our objective is to examine whether the tonal functionality traits exhibit consistent patterns or relationships within these curves. This exploration seeks to uncover more effective methods for modeling chord progressions in Bach's music, thereby advancing our understanding of the underlying structural principles and enhancing the authenticity of generated musical compositions.

Furthermore, we plan to explore deep-learning strategies to identify optimal interpolations (e.g., the Adversarially Constrained Autoencoder Interpolation and the Bridge Process [24, 25]). These methodologies, which have not yet been integrated into symbolic music VAEs, possess the potential to produce more authentic and stylistic interpolations. Their ability to discern interpolation patterns within the original chord progressions suggests an exciting avenue for future exploration. By incorporating these methodologies into symbolic music VAEs, we aim to enhance the fidelity and richness of generated musical compositions, ultimately advancing the state-of-the-art in computational music generation.

In summary, we plan to investigate the applicability of these sampling strategies to other latent spaces within the realm of music, such as audio latent spaces. By extending our exploration beyond symbolic representations, we aim to enhance sampling techniques and improve music generation across various types of models. This broader inquiry holds promise for advancing the field of computational music generation and facilitating the creation of more diverse and expressive musical compositions across different modalities.

Acknowledgments

This research has been funded by the Portuguese National Funding Agency for Science, Research and Technology [2021.05132.BD].

6. REFERENCES

[1] N. Carvalho and G. Bernardes, "Exploring latent spaces of tonal music using variational autoencoders," in *The International Conference on AI and Musical Creativity (AIMC)*, 2023.

[2] ——, "Fourier (common-tone) phase spaces are in tune with variational autoencoders' latent space," in *Mathematics and Computation in Music: 9th International Conference, MCM 2024, Coimbra, PT, June 18–21, Proceedings*, Springer-Verlag, Berlin, Heidelberg, 2024.

[3] N. Bryan-Kinns, "Reflections on explainable ai for the arts (xaixarts)," *Interactions*, vol. 31, no. 1, p. 43–47, jan 2024. [Online]. Available: https://doi.org/10.1145/3636457

[4] K. Chen, G. Xia, and S. Dubnov, "Continuous melody generation via disentangled short-term representations and structural conditions," in *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, 2020, pp. 128–135.

[5] A. Pati and A. Lerch, "Is disentanglement enough? on latent representations for controllable music generation," *Computing Research Repository (CoRR)*, vol. abs/2108.01450, 2021. [Online]. Available: https://arxiv.org/abs/2108.01450

[6] B. Banar, N. Bryan-Kinns, and S. Colton, "A tool for generating controllable variations of musical themes using variational autoencoders with latent space regularisation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 13, pp. 16 401–16 403, Sep. 2023. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/27059

[7] M. Turker, A. Dirik, and P. Yanardag, "Midispace: Finding linear directions in latent space for music generation," in *Proceedings of the 14th Conference on Creativity and Cognition*, ser. CC '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 420–427.

[8] Z. Guo, J. Kang, and D. Herremans, "A domain-knowledge-inspired music embedding space and a novel attention mechanism for symbolic music modeling," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, pp. 5070–5077, Jun. 2023. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/25635

[9] N. Bryan-Kinns, B. Zhang, S. Zhao, and B. Banar, "Exploring variational auto-encoder architectures, configurations, and datasets for generative music explainable ai," *Machine Intelligence Research*, vol. 21, no. 1, p. 29–45, Jan. 2024. [Online]. Available: http://dx.doi.org/10.1007/s11633-023-1457-1

[10] R. Yang, T. Chen, Y. Zhang, and G. Xia, "Inspecting and interacting with meaningful music representations

using VAE," *Computing Research Repository (CoRR)*, vol. abs/1904.08842, 2019. [Online]. Available: http://arxiv.org/abs/1904.08842

[11] R. Yang, D. Wang, Z. Wang, T. Chen, J. Jiang, and G. G. Xia, "Deep music analogy via latent representation disentanglement," in *20th International Society for Music Information Retrieval (ISMIR)*, 2019. [Online]. Available: https://archives.ismir.net/ismir2019/paper/000072.pdf

[12] A. Pati and A. Lerch, "Attribute-based regularization of latent spaces for variational auto-encoders," *Neural Computing and Applications*, vol. 33, no. 9, p. 4429–4444, Aug 2020.

[13] A. Pati, A. Lerch, and G. Hadjeres, "Learning to traverse latent spaces for musical score inpainting," in *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019, Delft, The Netherlands, November 4-8, 2019*, A. Flexer, G. Peeters, J. Urbano, and A. Volk, Eds., 2019, pp. 343–351. [Online]. Available: http://archives.ismir.net/ismir2019/paper/000040.pdf

[14] M. Prang and P. Esling, "Signal-domain representation of symbolic music for learning embedding spaces," *Computing Research Repository (CoRR)*, vol. abs/2109.03454, 2021. [Online]. Available: https://arxiv.org/abs/2109.03454

[15] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, "A hierarchical latent vector model for learning long-term structure in music," in *International Conference on Machine Learning*, vol. abs/1803.05428, 2018. [Online]. Available: http://arxiv.org/abs/1803.05428

[16] S. Ji, X. Yang, and J. Luo, "A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges," *ACM Comput. Surv.*, vol. 56, no. 1, aug 2023. [Online]. Available: https://doi.org/10.1145/3597493

[17] L. Mi, T. He, C. F. Park, H. Wang, Y. Wang, and N. Shavit, "Revisiting latent-space interpolation via a quantitative evaluation framework," *Computing Research Repository (CoRR)*, vol. abs/2110.06421, 2021. [Online]. Available: https://arxiv.org/abs/2110.06421

[18] B. V. Dasarathy, *Nearest neighbour norms*. Los Alamitos, CA: IEEE Computer Society Press, Dec. 1991.

[19] D. Huron, "Tone and voice: A derivation of the rules of voice-leading from perceptual principles," *Music Perception*, vol. 19, pp. 1–64, 09 2001.

[20] G. Dong, Y. Qingyun, Shang, Ye, and Y. Zhang, "Direction-aware continuous moving k-nearest-neighbor query in road networks," *ISPRS International Journal of Geo-Information*, vol. 8, p. 379, 08 2019.

[21] G. Bernardes, D. Cocharro, M. Caetano, C. Guedes, and M. E. Davies, "A multi-level tonal interval space for modelling pitch relatedness and musical consonance," *Journal of New Music Research*, vol. 45, no. 4, pp. 281–294, Oct. 2016, https://doi.org/10.1080/09298215.2016.1182192.

[22] G. Bernardes, D. Cocharro, C. Guedes, and M. E. Davies, "Harmony generation driven by a perceptually motivated tonal interval space," *Computers in Entertainment (CIE)*, vol. 14, no. 2, pp. 1–21, 2016.

[23] F. C. F. Almeida, G. Bernardes, and C. Weiß, "Mid-level harmonic audio features for musical style classification," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, 2022.

[24] D. Berthelot, C. Raffel, A. Roy, and I. J. Goodfellow, "Understanding and improving interpolation in autoencoders via an adversarial regularizer," *Computing Research Repository (CoRR)*, vol. abs/1807.07543, 2018. [Online]. Available: http://arxiv.org/abs/1807.07543

[25] C. Ringqvist, J. Butepage, H. Kjellström, and H. Hult, "Interpolation in auto encoders with bridge processes," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 5973–5980.