Information Synthesis of Time-Geometry QCurve for Music Retrieval

Shannon Steinmetz(✉) and Ellen Gethner

University of Colorado, Denver 80204, USA
{shannon.steinmetz,ellen.gethner}@ucdenver.edu

Abstract. We expand information segmentation to include additional properties of music geometry. We establish a distinct metric for invariant chord structure (harmonic consistency) and models for conjunct melodic motion and acoustic consonance. We combine these with centricity to form a unified measure of music geometry. Using geometric predictors and the LSQOP method, we classify music/non-music with comparable results to AI/ML, between 76% and 92% f-score.

Keywords: Music geometry · Information geometry · Harmonic consistency · Harmonic leading · Centricity · Dissonance · Consonance · Time-geometry · Qcurve · Quant-curve · Music retrieval · Music detection · LSQOP

1 Introduction

Music information retrieval (MIR) dominates audio classification, rhythm, melody, genre and emotion (MER) [19]. MIR began with self-similarity [8], but focuses now on neural networks (NN, CNN, RNN) [10] and support vector machines (SVM) [20]. Artificial intelligence (AI) and machine learning (ML) have proven successful with f-score as high as 85% [12,14]; however, AI/ML is burdened by data availability, supervision and labeling. This also means preprocessing (e.g. CUSUM, MLR, GLR, KCD [5,11]) is a major factor in finding valid segments to process.

We use Tymoczko's properties of music [23] as the basis for geometric audio segmentation. Therefore, we further develop sufficient models for these properties as segmentation estimators [22]. If we combine geometrically segmented (time/geometry) audio curves of the same class such as music, we arrive at *principal curves*, which are the foundation for testing geometric variance (i.e. classifying audio). Using a technique called LSQOP we will demonstrate classification for music and non-music, comparable to AI/ML [9,15].

S. Steinmetz—Work of the first author is supported in part by the Northrop Grumman Technical Fellowship.

E. Gethner—Work of the second author is supported in part by a Simons Foundation Collaboration Grant for Mathematicians.

M. Montiel et al. (Eds.): MCM 2022, LNAI 13267, pp. 398–404, 2022.
https://doi.org/10.1007/978-3-031-07015-0_36

2 QCurve Transform

Steinmetz and Gethner developed centricity segmentation via three parameter Gamma and geodesic likelihood [1,2,22]. The following sections expand this by modeling additional properties where the result is what we call time/geometry, or *qcurve.*

Definition 1. *(QCurve) A* qcurve *is a time series consisting of positive unitary measures of musical geometry.*

3 Harmonic Consistency

Harmonic consistency states "harmonies in a passage of music, whatever they may be, tend to be structurally similar to one another [23]."

Definition 2 (Musical Structure)**.** *Let* $K_{12} = (V, E)$, $V = \{0, 1, \cdots, 11\}$ *be a complete graph with vertices labeled* $\{0 = C, 1 = C\#, \cdots, 11 = B\}$*. Every distinct edge and cycle* C_r $3 \leq r \leq 12$ *having non-crossing edges in the embedded* K_{12} *is a* musical structure*.*

Definition 3. *Two non-empty sets of musical frequencies are* similar *if information gain due to geometric consistency is zero, or very small, such that distance* $D(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i+1)}) \leq \epsilon$ *where* ϵ *is a fixed positive value.* $D(\cdot)$ *is dependent on probabilities* $p(x_1 \,|\, \boldsymbol{\theta}^{(i)})$, $p(x_2 \,|\, \boldsymbol{\theta}^{(i+1)})$ *from [22].*

Definition 3 expands on Cont's idea of similarity [3] except here, we depend on geometric divergence. There are trivial, but useful mapping between vertices of a 12-gon and $\mathbb{Z}_{12}$ using the complex unit circle $z_s = f(s) = e^{i\pi(15-s)/6}$, $s \in \mathbb{Z}_{12}$. Assume musical frequency $k \in \mathbb{R}$, where $T : \mathbb{R} \rightarrow \mathbb{Z}$ such that $T(k) = \{0, 1, 2, \ldots, 11\}$ (i.e. pitch class). We define $\boldsymbol{S_t}(k_i, k_j)$ as shortest distance between Z_{12} elements, or *integer separation* between frequencies. Due to chroma, there exists a distinct, linearly independent, invariant Euclidian distance for every 12-tone musical frequency pair, therefore frequencies of identical integer separation are similar, which we will prove.

Because $s = f^{-1}(z) = \frac{6\arg(z)}{\pi}$ and $f^{-1}(a) = f^{-1}(b) \Rightarrow \frac{6\arg(f(a))}{\pi} = \frac{6\arg(f(b))}{\pi} \Rightarrow a = b$, f is injective. Let γ be inner angle difference between complex arguments. Due to injectivity, integer separation is modeled $\gamma = \gamma_i - \gamma_j = \frac{\boldsymbol{S}_t(k_1,k_2)\pi}{6}$.

Lemma 1. *Every pair of musical frequencies* k_i, k_j *has invariant Euclidian distance*

$$2\,sin\left(\frac{\boldsymbol{S_t}(k_i, k_j)\pi}{12}\right). \tag{1}$$

Proof. Let $z_i, z_j \in \mathbb{C}$, $z_i \neq z_j$. We define magnitude $|\cdot| \equiv ||\cdot||_2$. Due to the law of cosines $|z_i - z_j|^2 = |z_i|^2 + |z_j|^2 - 2\,|z_i||z_j|\,cos(\theta)$. Since $|z| = 1$, $|z_i - z_j|^2 = 2(1 - cos(\theta))$. If $\gamma_i = arg(z_i)$, $\gamma_j = arg(z_j)$, then by the dot product

$$\theta = cos^{-1}(cos(\gamma_i)cos(\gamma_j) + sin(\gamma_i)sin(\gamma_j)).$$

Substituting the angle and replacing identities gives $|z_i - z_j|^2 = 2(1 - cos(\gamma_i)cos(\gamma_j) + sin(\gamma_i)sin(\gamma_j))$. Multiply by $\frac{1}{4}$ and substitute haversine

$$\frac{|z_i - z_j|^2}{4} = \frac{1 - cos(\gamma_i - \gamma_j)}{2} \Rightarrow |z_i - z_j| = 2\,sin(\frac{\gamma_i - \gamma_j}{2}).$$

Constrained to the positive domain and substituting γ leaves

$$2\,sin\left(\frac{\boldsymbol{S_t}(k_i, k_j)\pi}{12}\right) = |z_i - z_j|.$$

Since rotation is unitary, all distinct pairs of musical frequencies have invariant Euclidian distance of this form. □

Lemma 1 is chroma-agnostic, linearly dependent, dyad similarity, but removing homogeneity makes all dyad pairs linearly independent under this mapping, therefore

$$\text{tone distance} = \delta_t(k_i, k_j) = 2\,sin\left(\frac{\boldsymbol{S_t}(k_i, k_j)\pi}{12}\right) + \boldsymbol{S_t}(k_i, k_j). \tag{2}$$

Tone distance is symmetric $\delta_t(k_i, k_j) = \delta_t(k_j, k_i)$, invariant under rotation and satisfies triangle inequality. Because $k_i = k_j$ implies $\delta_t(k_i, k_j) = 2 \cdot sin(0) + 0 = 0$, (2) is a *metric space*. It follows frequencies are similar if and only if, tone distance are equivalent. *Path distance* $= P_\delta(F_{(i)})$ is defined as the sum of tone distance, assuming frequencies $F_{(i)} = \{k_j\}_{j=1}^{n}$. Using logical reduction and brute force search, we find *no duplicate* path distances among all 12-tone chords. Temporally, inverse harmonic consistency $= \text{HC}^{-1} = \sigma(\boldsymbol{H}) \leq M_h = 18.24$, where σ is standard deviation, $\boldsymbol{H} = \{P_\delta(F_{(i)})\}_{i=1}^{\infty}$ and $F_{(i)}$ are sequential. Given no discernable notes, noise, or silence $\text{HC}^{-1} = M_h$.

4 Harmonic Leading

Harmonic leading is the combined measure of harmonic consistency, voice leading and conjunct melodic motion (CMM). CMM is the tendency for "melodies to move by short distances from note to note [23]." Due to [4] *musical metric distance* $\Delta(k_1, k_2)$ can be leveraged assuming $\boldsymbol{P} \in \mathbb{R}^{m \times n}$, whose columns are frequencies sorted in ascending order with $\boldsymbol{P}_{ij} = 0$, when m differs. From this, we approximate conjunct melodic motion

$$\overline{\nabla}(\boldsymbol{P}) = \left| \sum_{j=1}^{n-1} \sum_{i=1}^{m} \min\{\Delta(\boldsymbol{P}_{ij}, \boldsymbol{P}_{i(j+1)})\} \right|, \tag{3}$$

which also dampens overfit[1]. Assuming $\sigma(\cdot) = \text{stddev}$, $f = \text{frame count}$ and $\Delta(\cdot) \leq M_v = 6$ [4]

$$\text{harmonic leading} = 1 - \frac{\sigma(\boldsymbol{H}) + \overline{\nabla}(\mathbf{P})}{M_v(f-1) + M_h}. \tag{4}$$

[1] Harmonic leading *overfit* is defined as acoustically perceivable chord transposition, or inversion.

As a side note, we tested harmonic leading on Harte's 16 songs [13] with moderate success of 58% f-score, but we were unable to match HCDF retrieval. We recommend an HCDF quantifier experiment using our information framework, which we leave as an open problem.

5 Consonance

Definition 4. Acoustic consonance *is the inverse of dissonant contributions between* 20 *and* 250 *Hz of center frequency, within critical band* β *[7, 17, 18].*

Given $F(k) = |FFT|$, we select $\tau = |k - \text{argmax}_x \left[F(k) \cdot F(k+x)\right]|$, $x \leq 250$ which implies

$$\text{congruence score} = C_\tau(k) = \frac{1}{\max\{|F|\}} \sum_{x=k-\beta/2}^{k+\beta/2} F(x)\, f_\tau(x), \tag{5}$$

correlating $f_\tau(x) = \max\{F\}/2\left[cos(2\pi x/\tau) + 1\right]$. We center on $k_c : |k_c - k| \leq \beta/2$, selecting $k_j : F(k_j) \geq \frac{\max(F)}{4}$ and observe $\boldsymbol{C} = \left\{|C_{\tau_j}(k_j)|\right\}_{j=1}^{n}$ converge to a block wave as dissonant contribution increases. Therefore,

$$1 - \text{consonance} = \text{dissonance} = \frac{\max(\boldsymbol{C})}{2M_d} \sum_{j=1}^{n} D \cdot \boldsymbol{C}_j, \tag{6}$$

where max dissonance $= M_d \approx 10$ and $D = \text{sgn}\left[\cos(\frac{2\pi j}{s})\right] + 1$ over bandwidth $s \leq \beta$. β is displacement between the two loudest frequencies in critical band.

6 LSQOP

Qcurves are synthesized using [22] and combinations of $AC =$ acoustic consonance, $C =$ centricity, $HL =$ harmonic leading and $MG =$ music geometry.

$$MG = 1 - \frac{2}{\pi}\, cos^{-1}\left(\frac{U \cdot V}{\sqrt{3}\,||V||}\right) \tag{7}$$

is overlap between ideal $U = (1, 1, 1)$ and measured vector $V =$ (AC, C, HL). We observe ordinal separation between qcurves of differing classes, exposing an opportunity to train *principal qcurves* as a model for variation. Figure 1 illustrates least squares ortho projector (LSQOP) method, assuming input audio qcurve $q(x)$, principal qcurve trend-line segments $\{p_0, p_1\}$(music), $\{p_2, p_3\}$(non-music) and arbitrary point $d = (x, q(x))$. There exist $x_1 = p_0 + \text{Res}_{p_1 - p_0}(d - p_0)$ and $x_2 = p_1 + \text{Res}_{p_3 - p_2}(d - p_2)$ assuming Res$(\cdot)$ implies vector resolute. $\boldsymbol{P} = \frac{(0,1)(0,1)^T}{(0,1)^T(0,1)}$, $a_1 = ||\boldsymbol{P}d||$, $a_2 = ||\boldsymbol{P}x_1||$ and $a_3 = ||\boldsymbol{P}x_2||$, yields ratio

$$\mathfrak{m}(x) = \begin{cases} 1 & a_1 > a_2 \\ 0 & a_1 < a_3 \,\text{or}\, a_3 > a_2 \\ \frac{||a_1 - a_3||}{||a_2 - a_3||} & \text{otherwise.} \end{cases} \tag{8}$$

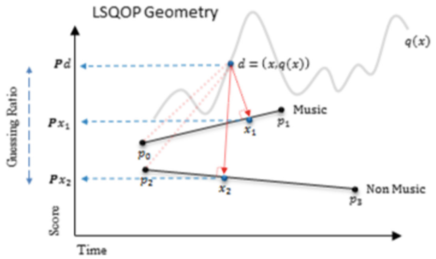

Fig. 1. Classification of audio qcurve $q(x)$, contrasted with trend lines from trained principal qcurves for music (top) and non-music (bottom).

Musical probability $\mathfrak{M} = E[m(x)]$ has binary pass/fail $\mathfrak{M} \geq \epsilon$ with expectation $E[\cdot]$ and non-negative *musical threshold* $\epsilon \leq 1$. Threshold varies depending on curve training.

7 Evaluation and Analysis

Verification of LSQOP and time/geometry involved classification experiments with GTZAN [24], TUT-17 (parts 1&2) [16], SWS1 (used by [6,21]), SWS2 (music), SWS3 (non-music) and SWS4 (non-music) data. The database contains 1556 music and 1441 non-music, totaling 2776 files. Custom data were created to fool LSQOP due to abnormal accuracy on GTZAN (100%) and TUT (92%). Custom data contains random quality, sample rate, content, size and non-thematic clips from *samplefocus.com*, *partnersinrhyme.com*, *bensound.com*, *freemp3cloud.com* and *BBC Sound Effects*. Custom non-music sets contain several categories (e.g. people, urban, construction, natural, office, animals, household, video games, military/war, etc.). Custom music sets are spread (mostly) even across several genres with famous, lesser known artists, synthesised and "poor" quality.

We performed three tests using LSQOP for music/non-music classification. The first and second test measures accuracy on all 2997 files. In the first we process the starting 10s of each file and in the second we process [.2, 3]s random samples from each file. Assuming notation (score/segment) (see [22]), (MG/HL) was effective with non-random samples at 96.4% accuracy for music and 70.78% classifying non-music. (MG/AC) was effective on random samples with 80% accuracy for music and 78% for non-music.

Table 1 is the result of an information retrieval (IR) exercise involving requests for music/non-music from a database (not all tests shown). Small randomly populated datasets are drawn from the entire clip database and each set is guaranteed to carry (roughly) equally distributed music and non-music. The results here are f-scores between 76% and 92%, averaging around 81.6%.

Table 1. Classification on randomly constructed subsets of the database. **M** = Music, **N** = Non-music, **n** = Num Random Samples, **p** = Precision, **r** = Recall, **f** = FScore. Training data shorthand: $A \equiv$ (SWS2/SWS4), $B \equiv$ (SWS2/TUT), $C \equiv$ GTZAN/SWS[1–3]. This table was inspired by [13].

Quantifier	Mn	Mp	Mr	Mf	Nn	Np	Nr	Nf
MG-C 10s	28	86%	92%	89%	28	93%	87%	90%
(AC/C)-B 10s	27	80%	92%	86%	27	92%	79%	85%
(MG/AC)-A 6s	22	100%	80%	89%	22	86%	100%	92%

8 Conclusions

We showed how 12-tone chroma-agnostic frequency pairs are mapped to a distinct, linearly independent, invariant measure of harmonic distance as a metric space. We provided a unique measure of musical chords (path distance) and discussed how to quantify harmonic consistency over successive frames. By combining this idea with work in voice leading we modeled conjunct melodic motion as harmonic leading. We then developed a straightforward approximation of acoustic consonance using psychoacoustic theory. From these measures we devised a unified score for musical geometry and showed how principal qcurves combined with LSQOP is effective in retrieval with consistent f-scores between 76% and 92%, averaging 81.6%.

Individually, the proposed musical property models have independent value for measuring structural content and acoustic quality analysis. The success of LSQOP is clear, but optimal time/geometry combinations require further experiments. Because audio data were successfully used to train accurate results against other audio, this opens the question of ideal data for use in widespread classification. We must also consider the disparity between music and non-music success, which itself can present a challenge to blind classification using the same geometric properties for music and non-music.

References

1. Abdel-All, N.H., Abdel-Galil, E.: Numerical treatment of geodesic differential. In: International Mathematical Forum, vol. 8, pp. 15–29 (2013)
2. Chen, W.W., Kotz, S.: The riemannian structure of the three-parameter gamma distribution (2013)
3. Cont, A., Dubnov, S., Assayag, G.: On the information geometry of audio streams with applications to similarity computing. IEEE Trans. Audio Speech Lang. Process. **19**, 837–846 (2010)
4. del Pozo, I., Gómez, F.: Formalization of voice-leadings and the Nabla algorithm. In: Montiel, M., Gomez-Martin, F., Agustín-Aquino, O.A. (eds.) MCM 2019. LNCS (LNAI), vol. 11502, pp. 352–358. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21392-3_30

5. Desobry, F., Davy, M., Doncarli, C.: An online kernel change detection algorithm. IEEE Trans. Signal Process. **53**, 2961–2974 (2005)
6. Gethner, S.S.E., Verbeke, J.: A view of music. In: Delp, D.M.K., Kaplan, C.S., Sarhangi, R. (eds.) Proceedings of Bridges 2015: Mathematics, Music, Art, Architecture, Culture, Phoenix, Arizona, Tessellations Publishing, pp. 289–294 (2015). http://archive.bridgesmathart.org/2015/bridges2015-289.html
7. Fishman, Y.I., et al.: Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J. Neurophysiol. **86**, 2761–2788 (2001)
8. Foote, J.T., Cooper, M.L.: Media segmentation using self-similarity decomposition. In: Storage and Retrieval for Media Databases, vol. 5021, pp. 167–175. International Society for Optics and Photonics (2003)
9. Gimeno, P., Mingote, V., Giménez, A.O., Miguel, A., Lleida, E.: Partial auc optimisation using recurrent neural networks for music detection with limited training data. In: Interspeech, pp. 3067–3071 (2020)
10. Gururani, S., Summers, C., Lerch, A.: Instrument activity detection in polyphonic music using deep neural networks. In: ISMIR, pp. 569–576 (2018)
11. Gustafsson, F.: The marginalized likelihood ratio test for detecting abrupt changes. IEEE Trans. Autom. Control **41**, 66–78 (1996)
12. Hamel, P., Eck, D.: Learning features from music audio with deep belief networks. In: ISMIR, vol. 10, pp. 339–344. Citeseer (2010)
13. Harte, C., Sandler, M., Gasser, M.: Detecting harmonic change in musical audio. In: Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, pp. 21–26 (2006)
14. Kataoka, M., Kinouchi, M., Hagiwara, M.: Music information retrieval system using complex-valued recurrent neural networks. In: SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218), vol. 5, pp. 4290–4295. IEEE (1998)
15. Meléndez-Catalán, B., Molina, E., Gomez, E.: Music and/or speech detection mirex 2018 submission, music information retrieval evaluation eX-change (2018)
16. Mesaros, A., Heittola, T., Virtanen, T.: Tut acoustic scenes 2017, evaluation dataset (2017)
17. Moore, B.C., Glasberg, B.R.: Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. **74**, 750–753 (1983)
18. Nordmark, J., Fahlén, L.E.: Beat theories of musical consonance. Q. Prog. Status Rep. **29**, 111–122 (1988)
19. Panda, R., Malheiro, R.M., Paiva, R.P.: Audio features for music emotion recognition: a survey, IEEE Trans. Affect. Comput. (2020)
20. Schedl, M., Gutiérrez, E.G., Urbano, J.: Music information retrieval: recent developments and applications. Found. Trends Inf. Retrieval. **8**(2–3), 127–261 (2014)
21. Steinmetz, S., Sonic imagery: a view of music via mathematical computer science and signal processing. University of Colorado at Denver (2016)
22. Steinmetz, S., Gethner, E.: On musical information geometry with applications to sonified image analysis. In: To Appear in International Conference on Pattern Recognition and Image Processing, Miami (2021)
23. Tymoczko, D.: A Geometry of Music, Oxford University Press, Oxford, 1 (ed.) (2011)
24. Zhang, X.: Gtzan, November 2019