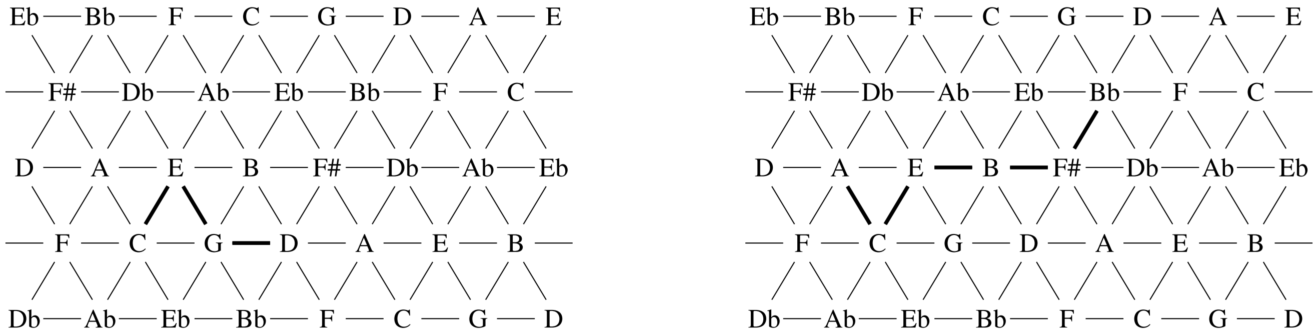

(a) The shortest path passing through C, D, E, and G.

(b) The shortest path passing through C, F, A, and Bb.

Figure 1: Calculating the `ChordTonnetz` feature for different pitch class sets

A FEATURES

In what follows, we provide additional explanations for some of the features described in Table 1. The function `sc` calculates the scale representation described in the text. Let $\mathbb{S}^{\texttt{maj}} = \{0, 2, 4, 5, 7, 9, 11\}$ and $\mathbb{S}^{\texttt{harm}} = \{0, 2, 3, 5, 7, 8, 11\}$, then $\mathbb{S}_i^{\texttt{maj}} = \{(s+i) \mod 12 : s \in \mathbb{S}^{\texttt{maj}}\}$ denotes the major scale with pitch class i as a root, and $\mathbb{S}_i^{\texttt{harm}} = \{(s+i) \mod 12 : s \in \mathbb{S}^{\texttt{harm}}\}$ denotes the harmonic minor scale with pitch class i as a root. The scale representation can be calculated using Equation 6, where x is a pitch class set, and $\phi(\cdot)$ is a function that returns 1 if the predicate $\cdot$ is true, and 0 otherwise.

$$\texttt{sc}(x) = \big(\sum_{i=1}^{12} \phi(x \subseteq \mathbb{S}_i^{\texttt{maj}}) \ll i\big) + \big(\sum_{i=1}^{12} \phi(x \subseteq \mathbb{S}_i^{\texttt{harm}}) \ll (12+i)\big) \tag{6}$$

The difference between two pitches p_1 and p_2 can be represented in several ways: as the absolute difference between two pitches ($d_{\texttt{abs}}(p_1, p_2) = |p_1 - p_2|$); as an interval ($d_{\texttt{mod}}(p_1, p_2) = (p_1 - p_2) \mod 12$); and as an interval class. There are 6 interval classes: the unison ($d_{\texttt{mod}}(p_1, p_2) \in \{0\}$; the minor second and major seventh ($d_{\texttt{mod}}(p_1, p_2) \in \{1, 11\}$; the major second and minor seventh ($d_{\texttt{mod}}(p_1, p_2) \in \{2, 10\}$; the minor third and major sixth ($d_{\texttt{mod}}(p_1, p_2) \in \{3, 9\}$; the major third and minor sixth ($d_{\texttt{mod}}(p_1, p_2) \in \{4, 8\}$; the perfect fourth and fifth ($d_{\texttt{mod}}(p_1, p_2) \in \{5, 7\}$; and the tritone ($d_{\texttt{mod}}(p_1, p_2) \in \{6\}$. We calculate the interval class using Equation 7,

$$\texttt{pcc}(p_1, p_2) = |((p_1 - p_2) \mod 12) - 6| \tag{7}$$

The `ChordTonnetz` feature is simply the length of the shortest path which passes through each of the pitch classes contained in a chord. For example, given a chord $\mathbb{C}$ and the corresponding pitch set $\mathbb{P} = \{48, 55, 60, 62, 64\}$, the shortest path passing though each pitch class in the set {C,D,E,G} is shown in Figure 1a. In this case, the length of the shortest path is 3. In general, harmonically simple chords will have shorter paths than harmonically complex chords. The shortest path for the pitch class set {C,F#,A,Bb} is shown in Figure 1b.

The `ChordTranVoiceMotion` feature denotes the type of voice motion between two successive chords. Given two successive pitch class sets $\mathbb{P}^t$ and $\mathbb{P}^{t+1}$ Equation 8 is used to calculate the type of voice motion, where 0 is no change, 1 is oblique motion, 2 is parallel motion, and 3 is contrary motion.

$$\texttt{ChordTranVoiceMotion}(\mathbb{P}^t, \mathbb{P}^{t+1}) = \begin{cases} 0, & \text{if } (\min(\mathbb{P}^t) = \min(\mathbb{P}^{t+1})) \wedge (\max(\mathbb{P}^t) = \max(\mathbb{P}^{t+1})) \\ 1, & \text{if } (\min(\mathbb{P}^t) = \min(\mathbb{P}^{t+1})) \wedge (\max(\mathbb{P}^t) \neq \max(\mathbb{P}^{t+1})) \\ 1, & \text{if } (\min(\mathbb{P}^t) \neq \min(\mathbb{P}^{t+1})) \wedge (\max(\mathbb{P}^t) = \max(\mathbb{P}^{t+1})) \\ 2, & \text{if } (\min(\mathbb{P}^t) > \min(\mathbb{P}^{t+1})) \wedge (\max(\mathbb{P}^t) > \max(\mathbb{P}^{t+1})) \\ 2, & \text{if } (\min(\mathbb{P}^t) < \min(\mathbb{P}^{t+1})) \wedge (\max(\mathbb{P}^t) < \max(\mathbb{P}^{t+1})) \\ 3, & \text{otherwise} \end{cases} \tag{8}$$

Here, we briefly describe the method for calculating chord periodicity, which is used to calculate the `ChordDissonance` and `ChordTranDissonance` features. For more specific details please consult the original paper [32]. Given a set of pitches $\mathbb{P}_0 = \{0, 3, 9\}$, which corresponds to the frequency ratios $\mathbb{F}_0 = \{1/1, 6/5, 5/3\}$, the lowest common multiple of the denominators is calculated ($L_0 = \text{lcm}(1, 5, 3) = 15$). To calculate the smoothed periodicity, L_1 and L_2 are calculated using the shifted pitch sets $\mathbb{P}_1 = \{-3, 0, 6\}$ and $\mathbb{P}_2 = \{-9, -6, 0\}$. The L_i values are scaled, and the average of the three values is taken. `ChordTranDissonance` slighlty modifies the above procedure. Given two pitch sets $\mathbb{P}^t = \{0, 3, 9\}$ and $\mathbb{P}^{t+1} = \{0, 4, 7\}$, the L values are calculated for the pitch sets $\mathbb{P}_0 = \{0, 4, 7\}$, $\mathbb{P}_1 = \{-3, 1, 4\}$, $\mathbb{P}_2 = \{-9, -5, -2\}$. We construct these shifted pitch sets as follows, where $\mathbb{P}_i = \{p - \mathbb{P}^t_i : p \in \mathbb{P}^{t+1}\}$. The same scaling procedure is applied here, and the average of the L_i values is calculated. Since `ChordDissonance` and `ChordTranDissonance` must be turned into integers, we use the floor operator to turn a float into an integer. In order to have a larger number of categories, we do not apply the log transform to each L_i value.

B EXPERIMENT 1 EXPANDED

In order to demonstrate that StyleRank is robust when the size of $\mathcal{G}$, and the number of styles in $\mathcal{G}$ are varied, we modify Experiment 1. Given $2 + k$ styles $S^i = \{s_1^i, ..., s_{m_j}^i\}$, where $m_1 = 2n$, $m_j = n$ for $j > 1$, let $\mathcal{C} = \{s_i^1 : n < i \leqslant 2n\}$, $\mathcal{G}_A = \{s_i^1 : 0 \leqslant i < n\}$, $\mathcal{G}_B = \{s_i^2 : 0 \leqslant i < n\}$, $\mathcal{G} = \{s_i^j : (0 \leqslant i < n) \wedge (1 \leqslant j \leqslant 2 + k)\}$. We train a random forest and compare two distributions $x = [S_g^{\mathcal{G},\mathcal{C},\mathcal{F}} : g \in \mathcal{G}_A]$ and $y = [S_g^{\mathcal{G},\mathcal{C},\mathcal{F}} : g \in \mathcal{G}_B]$. When $k = 0$, the process is identical to Experiment 1. The results for $0 \leqslant k < 4$ are shown in Table 4, with the top scores bolded for each combination of corpus size (n) and style type (`genre`, `composer`).

On a whole, the performance does not decrease as k increases, evidenced by the fact that $k > 0$ models had better scores than $k = 0$ in many cases. In the most extreme case, with $k = 3$ and $n = 100$, the size of $\mathcal{C}$ and $\mathcal{G}$ vary significantly, where $||\mathcal{C}|| = 100$ and $||\mathcal{G}|| = 500$. We believe this provides compelling evidence that StyleRank is robust against discrepancies in size between $\mathcal{C}$ and $\mathcal{G}$, and variations to the number of styles in $\mathcal{G}$.

			StyleRank				Cosine				Manhattan				Euclidean			
	k	size	μ	Sig	FDR	Bon	μ	Sig	FDR	Bon	μ	Sig	FDR	Bon	μ	Sig	FDR	Bon
Genre	0	10	**0.81**	**0.379**	0.0	0.0	0.68	0.193	0.0	0.0	0.686	0.2	0.0	0.0	0.645	0.176	0.0	0.0
	1	10	0.802	0.355	0.0	0.0	0.689	0.183	0.0	0.0	0.672	0.194	0.0	0.0	0.633	0.164	0.0	0.0
	2	10	0.764	0.332	0.0	0.0	0.695	0.212	0.0	0.0	0.68	0.189	0.0	0.0	0.647	0.155	0.0	0.0
	3	10	0.744	0.314	0.0	0.0	0.713	0.209	0.0	0.0	0.687	0.209	0.0	0.0	0.651	0.168	0.0	0.0
	0	25	**0.867**	0.578	**0.376**	**0.198**	0.729	0.348	0.084	0.038	0.74	0.374	0.053	0.021	0.691	0.298	0.06	0.022
	1	25	0.858	0.567	0.339	0.181	0.714	0.343	0.084	0.034	0.752	0.361	0.027	0.012	0.674	0.276	0.037	0.017
	2	25	0.841	**0.592**	0.357	0.197	0.73	0.359	0.068	0.019	0.743	0.363	0.048	0.016	0.71	0.289	0.038	0.017
	3	25	0.846	0.581	0.352	0.179	0.736	0.338	0.068	0.024	0.722	0.363	0.041	0.011	0.696	0.286	0.033	0.009
	0	50	0.88	0.715	0.59	0.432	0.776	0.484	0.266	0.126	0.747	0.489	0.253	0.088	0.714	0.344	0.158	0.082
	1	50	0.897	**0.748**	**0.621**	**0.458**	0.731	0.442	0.236	0.126	0.773	0.488	0.258	0.088	0.744	0.375	0.19	0.117
	2	50	**0.901**	0.739	0.608	0.437	0.755	0.449	0.239	0.118	0.74	0.492	0.287	0.101	0.728	0.366	0.18	0.109
	3	50	0.892	0.732	0.577	0.409	0.772	0.451	0.22	0.105	0.752	0.474	0.273	0.097	0.738	0.372	0.172	0.103
	0	100	0.927	0.847	0.774	0.671	0.766	0.555	0.406	0.265	0.755	0.566	0.44	0.284	0.785	0.462	0.269	0.178
	1	100	**0.941**	0.855	0.794	0.679	0.768	0.541	0.406	0.279	0.741	0.541	0.416	0.291	0.783	0.439	0.265	0.189
	2	100	0.93	0.835	0.775	0.668	0.77	0.562	0.395	0.26	0.753	0.547	0.446	0.308	0.777	0.457	0.264	0.192
	3	100	0.937	**0.865**	**0.802**	**0.697**	0.758	0.551	0.4	0.277	0.732	0.564	0.458	0.302	0.785	0.468	0.274	0.199
Composer	0	10	0.963	**0.86**	**0.725**	0.0	0.837	0.624	0.381	0.0	0.879	0.662	0.413	0.0	0.827	0.565	0.28	0.0
	1	10	**0.971**	0.849	0.704	0.0	0.85	0.626	0.401	0.0	0.862	0.628	0.392	0.0	0.812	0.568	0.292	0.0
	2	10	0.96	0.838	0.646	0.0	0.873	0.633	0.392	0.0	0.843	0.631	0.369	0.0	0.83	0.579	0.314	0.0
	3	10	0.954	0.818	0.657	0.0	0.857	0.635	0.388	0.0	0.876	0.663	0.437	0.0	0.835	0.576	0.288	0.0
	0	25	0.951	**0.888**	0.807	0.609	0.808	0.583	0.422	0.24	0.793	0.578	0.415	0.244	0.729	0.532	0.363	0.226
	1	25	0.946	0.879	**0.812**	**0.613**	0.773	0.561	0.403	0.246	0.818	0.627	0.461	0.276	0.738	0.533	0.36	0.196
	2	25	0.941	0.88	0.808	**0.613**	0.782	0.58	0.422	0.248	0.797	0.601	0.424	0.221	0.746	0.546	0.38	0.219
	3	25	**0.952**	0.881	0.796	0.61	0.802	0.57	0.422	0.246	0.798	0.63	0.431	0.247	0.764	0.568	0.377	0.219
	0	50	0.926	0.905	0.873	0.78	0.705	0.559	0.454	0.333	0.751	0.599	0.468	0.34	0.717	0.565	0.428	0.3
	1	50	**0.943**	**0.917**	0.885	0.785	0.722	0.567	0.444	0.308	0.697	0.54	0.435	0.305	0.686	0.531	0.369	0.262
	2	50	0.936	0.902	0.869	0.784	0.732	0.574	0.465	0.326	0.743	0.578	0.453	0.318	0.693	0.549	0.437	0.312
	3	50	0.942	**0.917**	**0.892**	**0.787**	0.714	0.557	0.448	0.316	0.721	0.571	0.454	0.325	0.676	0.54	0.421	0.306
	0	100	**1.0**	0.986	0.973	**0.951**	0.713	0.636	0.59	0.515	0.723	0.633	0.568	0.486	0.715	0.626	0.571	0.504
	1	100	**1.0**	**0.988**	**0.978**	0.941	0.745	0.662	0.604	0.539	0.724	0.636	0.572	0.492	0.712	0.62	0.552	0.489
	2	100	0.997	0.986	0.97	0.929	0.709	0.623	0.577	0.52	0.731	0.645	0.579	0.5	0.7	0.63	0.572	0.504
	3	100	0.994	0.975	0.961	0.918	0.71	0.631	0.585	0.505	0.753	0.665	0.606	0.523	0.689	0.61	0.551	0.494

Table 4: The normalized frequency over 1000 trials where $\bar{x} > \bar{y}$ (μ), $p^{\bar{x}>\bar{y}} < 0.05$ (Sig), $p^{\bar{x}>\bar{y}}$ is significant after applying the FDR correction (FDR), and $p^{\bar{x}>\bar{y}}$ is significant after applying the Bonferonni correction (Bon). Size denotes the size of the corpus $||\mathcal{C}|| = ||\mathcal{G}_k|| = ||\mathcal{G}_B||$.

C EXPERIMENT 1 DATA

The number of pieces belonging to each genre and composer after duplicates have been removed are shown in Table 5 and 6 respectively. Only the composers with more than $2n$ pieces are selected for comparison. As a result, there are only 8 composers to compare when $n = 100$.

Genre	Count
Middle Romantic	447
Post Romantic	515
Late Romantic	577
Early Romantic	777
Late Classical Early Romantic	1100
Late Baroque	2615

Table 5: The number of pieces in each genre.

Composer	Count	Composer	Count	Composer	Count
Johann Sebastian Bach	1081	Johann Friedrich Burgmüller	56	Sir Arthur Sullivan	30
Wolfgang Amadeus Mozart	813	Henry Purcell	54	Michael Maier	30
Domenico Scarlatti	551	Georg Philipp Telemann	54	Thomas Morley	29
George Frideric Handel	482	Sergey Vasilyevich Rachmaninov	52	Johann Adolf Hasse	28
Ludwig Van Beethoven	384	Maurice Ravel	52	Bedrich Smetana	27
Franz Liszt	309	Domenico Zipoli	48	Stephen Heller	26
(franz) Joseph Haydn	308	Muzio Clementi	45	Nikolay Rimsky-korsakov	26
Antonio Vivaldi	212	Igor Stravinsky	44	Jean-philippe Rameau	26
Frédéric François Chopin	197	Carl Maria Von Weber	44	Jean-baptiste Lully	26
Franz Peter Schubert	197	Niccolò Paganini	42	Giovanni Pierluigi Da Palestrina	26
Johannes Brahms	184	Jose Mauricio Nunes García	39	Scott Joplin	25
Felix Mendelssohn-bartholdy	141	Erik Satie	39	Leopold Godowsky	25
Johann Nepomuk Hummel	135	Charles-valentin Alkan	39	Gustav Mahler	25
Pyotr Ilyich Tchaikovsky	125	Silvius Leopold Weiss	37	Karl Joachim Andersen	24
Antonín (leopold) Dvořák	125	John Philip Sousa	37	Jacques Offenbach	24
Robert Alexander Schumann	117	John Dowland	37	Anton Bruckner	24
Achille-claude Debussy	103	(wilhelm) Richard Wagner	37	Sir Edward Elgar	23
William Byrd	92	Mauro Giuliani	35	Giovanni Battista Pergolesi	23
Carl Czerny	81	Marin Marais	35	Lorenzo Perosi	22
Camille Saint-saëns	81	Giovanni Battista Sammartini	35	Jean-baptiste Lemire	22
Alexander Scriabin	77	Charles Gounod	35	Giacomo Puccini	22
Isaac Albéniz	68	Béla Bartók	35	Richard Walthew	21
Fernando Sor	66	Johann Pachelbel	34	Ferdinando Carulli	21
Gabriel Fauré	64	Georges Bizet	34	Adriano Banchieri	21
Edvard Grieg	62	Modest Petrovich Mussorgsky	33	Gaetano Donizetti	20

Table 6: The number of pieces per composer.

D EXPERIMENT 2 DATA

In Table 7, we provide the raw frequency counts for each piece, from which our ground truth ranking is constructed. We also calculated the percentage of pairwise comparisons that were identical for different participant levels (Novice, Intermediate, Advanced and Expert), shown in Table 8. The ranking constructed from Novice data is the most dissimilar from the other three levels. For varying levels of α, we show the number of significant comparisons for each level in Table 9.

	Novice		Intermediate		Advanced		Expert	
	N^{corr}	N^{miss}	N^{corr}	N^{miss}	N^{corr}	N^{miss}	N^{corr}	N^{miss}
BWV-310-mask-Alto-Tenor-fermatas	113	53	308	70	142	26	53	3
BWV-378-mask-Alto-Tenor-Bass-fermatas	121	57	267	66	146	31	73	6
BWV-11.6-mask-Alto-Tenor-fermatas	107	71	234	88	124	38	49	6
BWV-419-mask-Soprano-fermatas	129	39	289	69	142	28	68	9
BWV-430-mask-Alto-Tenor-fermatas	116	60	290	63	142	17	61	9
BWV-411-mask-Alto-Tenor-fermatas	107	73	256	96	128	28	53	8
BWV-121.6-mask-Soprano-fermatas	108	68	254	87	159	39	63	11
BWV-276-mask-Alto-Tenor-Bass-fermatas	120	65	245	113	146	34	48	10
BWV-372-mask-Alto-Tenor-Bass-fermatas	97	59	269	78	130	35	52	12
BWV-127.5-mask-Alto-Tenor-fermatas	74	72	233	104	125	42	59	14
BWV-381-mask-Alto-Tenor-fermatas	112	60	286	74	143	24	45	11
BWV-166.6-mask-Alto-Tenor-fermatas	94	69	296	77	151	32	53	14
BWV-65.2-mask-Alto-Tenor-fermatas	91	57	205	106	130	49	57	16
BWV-268-mask-Alto-fermatas	92	59	247	108	126	38	54	16
out-28	124	100	305	186	139	62	60	19
BWV-154.3-mask-Alto-Tenor-Bass-fermatas	97	68	219	138	125	50	52	17
out-45	127	108	283	207	154	80	66	23
out-59	122	103	280	174	160	72	53	19
BWV-168.6-mask-Alto-Tenor-Bass-fermatas	89	74	221	103	113	36	38	14
BWV-425-mask-Alto-Tenor-Bass-fermatas	96	57	266	83	127	32	56	21
out-54	116	100	311	158	151	66	68	26
out-56	102	113	289	167	165	44	52	21
out-19	138	92	264	171	164	65	66	27
out-20	108	106	253	175	136	93	47	24
BWV-438-mask-Alto-fermatas	107	75	204	152	111	50	41	22
BWV-270-mask-Alto-Tenor-Bass-fermatas	89	77	213	153	88	78	33	18
out-33	136	96	290	191	131	99	57	32
BWV-114.7-mask-Tenor-fermatas	79	74	198	147	104	64	47	27
out-60	93	129	224	242	118	97	55	37
BWV-248.5-mask-Alto-Tenor-Bass-fermatas	90	73	177	175	91	79	36	25
out-63	99	116	233	261	113	108	51	43
BWV-102.7-mask-Tenor-fermatas	88	75	191	154	89	60	21	19
BWV-27.6-mask-Bass-fermatas	91	71	186	174	89	77	34	31
BWV-293-mask-Bass-fermatas	77	97	166	160	78	87	24	24
out-10	131	93	222	208	109	129	39	51
out-52	125	81	235	243	84	143	36	50
total	3805	2840	8909	5021	4573	2132	1820	735

Table 7: Raw count data from the BachBot experiment [19], where N^{miss} is the number of times a generated sample was mistakenly classified as a Bach chorale and N^{corr} is the number of times it was correctly identified as computer generated.

	Novice	Intermediate	Advanced	Expert
Novice	1.0	0.765	0.765	0.747
Intermediate	0.765	1.0	0.904	0.873
Advanced	0.765	0.904	1.0	0.870
Expert	0.747	0.873	0.870	1.0

Table 8: Percentage of identical pairwise comparisons, based on data from the Bachbot experiment [19].

	$\alpha = 5.0$	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.005$
Novice	630	432	174	82
Intermediate	630	554	429	351
Advanced	630	533	385	304
Expert	630	484	262	150

Table 9: Number of comparisons below significance level, based on data from the Bachbot experiment [19].