µTunes: A Study of Musicality Perception in an Evolutionary Context

Article · January 2014

4 authors, including:

Some of the authors of this publication are also working on these related projects:

Project   visual speech animation View project

Project   EU H2020 SWITCH (Software Workbench for Interactive, Time Critical and Highly self-adaptive Cloud applications) View project

μTunes: A Study of Musicality Perception in an Evolutionary Context

Kirill Sidorov   **Robin Hawkins**   **Andrew Jones**   **David Marshall**

Cardiff University, UK

K.Sidorov@cs.cardiff.ac.uk   ontario.cs.cf.ac.uk/mutunes

ABSTRACT

We have conducted an experiment with the intent to determine and quantify what properties of monophonic melodies humans perceive as appealing. This was done in an evolutionary setting: a population of melodies was subjected to Darwinian selection with popular human vote serving as the basis for the fitness function. We describe the experimental procedure, measures to avoid or minimise possible experimental biases, and address the problem of extracting maximum fitness information from sparse measurements.

We have rigorously analysed the course of the resulting evolutionary process and have identified several important trends. In particular, we have observed a decline in complexity of melodies over time, increase in diatonicity, consonance, and rhythmic variety, well-defined principal directions of evolution, and even rudimentary evidence of speciation and genre-forming. We discuss the relevance of these effects to the question of what is perceived as a pleasant melody.

Such analysis has not been done before and hence the novel contribution of this paper is the study of the psychological biases and preferences when popular vote is used as the fitness function in an evolutionary process.

1. INTRODUCTION

Evolutionary approach to music composition is well described in the literature: whether the fitness information is provided by human evaluation [1, 2] or otherwise [3, 4].

Recently, the importance of consumers' preference in driving the evolution of music was demonstrated in [2]. While their conclusions were criticised (especially as to the role of biases in selection [5]), the experiment of [2] was the first large-scale attempt at music evolution with popular vote serving as the fitness function. Further, [5, 6, 7] argue that recombination and transformation of information according to psychological *biases* of individuals are the crucial element of cultural evolution.

In contrast to [2], where the process of evolution itself was examined, this work concentrates on and attempts to measure the above-mentioned psychological and cultural biases that guide the evolution of music, and hence attempts to quantify what aspects of music the respondents find appealing.

2. EXPERIMENTAL SETUP

In this experiment, we maintain and evolve a population of melodies. To minimise the model bias, we have adopted the simplest representation of the population in which the phenotypes and genotypes of individuals are identically equivalent. Each individual is represented by two lists: a list of intervals (in semitones) between successive notes in the melody, and a list of note durations (as integer multiplies of the time quantum, in this case $\Delta t = 1/16$th note). The total duration of each melody is capped at $64 \times 1/16$th, which is equivalent to four bars in $4/4$.

At the start of evolution, the population is initialised with randomly generated melodies in which the intervals are drawn from integer Gaussian distribution with $\mu = 0$ and $\sigma = 7$ semitones (middle C is chosen for the first note in all melodies). Gaussian sampling is done to avoid biasing the respondents towards diatonicity. The note durations in the initial population are chosen equal (crotchets).

Thus, the initial melodies are results of Gaussian random walks, in which the expected root-mean-square deviation of the last note after $n = 15$ steps is $\sigma\sqrt{n} \approx 27$ semitones. The melodies were then confined (by taking the modulo) to the range of ± 1 octave from the starting note.

The population size is kept constant at every generation ($N = 100$ exemplars), with the entire evolutionary history being recorded for future analysis. The choice of the population size depends on two factors. First, the population size should ideally be large enough for emergent phenomena, such as speciation, to be observed. Second, too large a population would not allow us to observe a substantial number of generations: indeed, if ranking the population involves $O(N \log N)$ comparisons, then given a budget of C comparisons the number of generations we could observe is at most $\lfloor C/O(N \log N) \rfloor$.

To sample popular opinion, we set up a website [1] which offers the visitors two melodies from the current population. The visitors are prompted to play back the melodies and select the one which they prefer. Their response is recorded and serves to update the population rank.

In contrast to [2], where respondents choose between five categories, in our experiment they were presented with two options (select the best from two melodies). The reason for

---

[1] http://ontario.cs.cf.ac.uk/mutunes
The name "μTunes" is pronounced to rhyme with "mutants".

this is twofold. First, the problem of optimal ranking using expensive pairwise comparisons is well-studied in the literature [8, 9]. This is a non-trivial task, since we need to find the best estimate of the population rank given the limited number of pairwise comparisons which, in addition, may be noisy (contradictory). In [9] an efficient ranking algorithm is described which by adaptive pairwise queries efficiently ranks the population. Due to its robustness under noise this is the algorithm we used in our experiment.

After each comparison, with probability $N\gamma \log(N\gamma)$ a new generation is produced. Above, $N = 100$ is the population size and $\gamma = 0.5$ is a parameter controlling the rate of evolution. This is equivalent to triggering a new generation on average every $O(N \log N)$ comparisons which are required to rank N individuals.

When a new generation is triggered, the highest ranked $\alpha = 20\%$ of the individuals take part in sexual reproduction. Pairs are formed by uniform random sampling from the top $\alpha = 20\%$ of the population. Breeding involves a one-point crossover operation: a time t_f in the female melody is selected uniformly randomly (at any point, including in the middle of notes), with t_f being quantised to $1/16$th notes; similarly t_m is selected for the male individual. The melodies (intervals and note durations) are sliced at t_f and t_m.

Breeding produces two offsprings: one shares a beginning with the female parent, and an end with the male parent, and vice-versa for the other offspring. Such crossover can result in a note being broken in two, as slicing time t is not guaranteed to coincide with the note boundaries. When this occurs, with probability $\beta = 50\%$ the notes at the cut-point are fused. This way, we ensure that the rhythmic granularity does not unnecessarily increase due to crossover. A better approach would be to estimate β dynamically to explicitly ensure that the crossover operation does not affect the average granularity in the population.

Finally, mutation occurs in one of the offspring. The offspring to be mutated is chosen uniformly randomly. Mutation affects only the intervals. One interval, selected again uniformly at random, is incremented or decremented by one. The bottom $\alpha = 20\%$ of the population are removed from the population and replaced with the newly generated offspring. The rhythm was not explicitly subjected to mutation to simplify the experiment; however, variability in rhythm naturally arises as the result of the crossover operation, which splices the melodies as described above.

3. ANALYSIS AND RESULTS

So far into the experiment, we have registered $\approx 7,000$ comparisons which resulted in 45 generations of evolution. We have also carried out a control experiment in which the evolution proceeded under the same conditions (and the same initial population) except the responses were replaced with random Bernoulli-distributed values ($p = 1/2$, fair coin).

Below we describe the various features of the population that we measured over time as the evolution progresses: entropy of melody and rhythm, repetitiveness, properties of melodic contour, and pitch distribution.
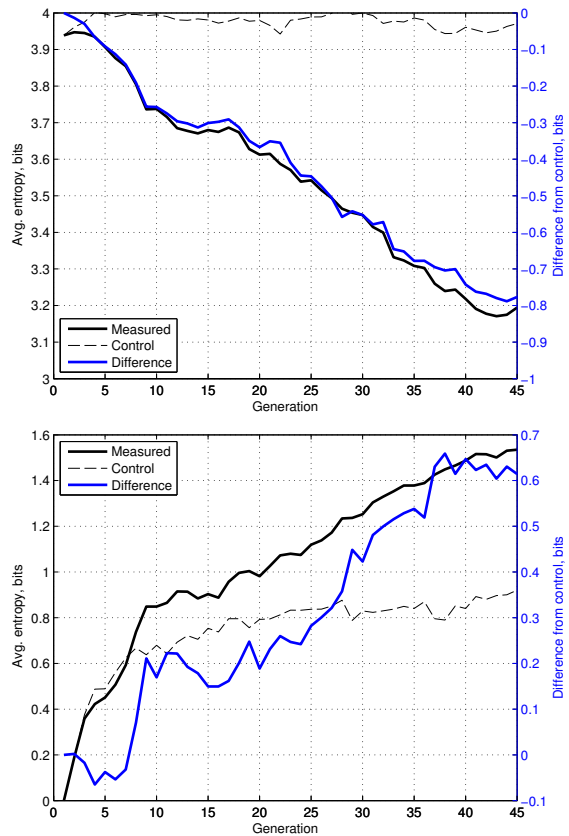

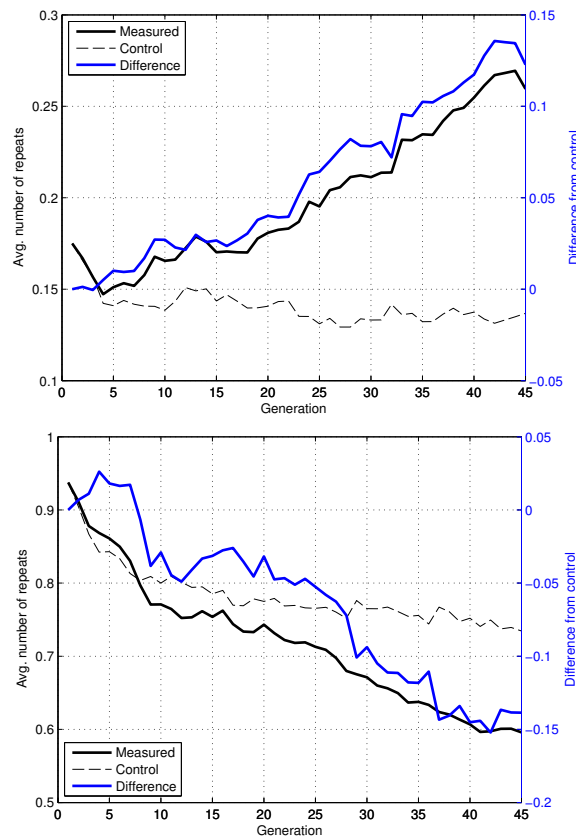**Figure 1**. Change in average entropy of melody (*above*) and rhythm (*below*).

Figure 2. Average number of consecutive repetitions of pitches (*above*) and rhythmic values (*below*).
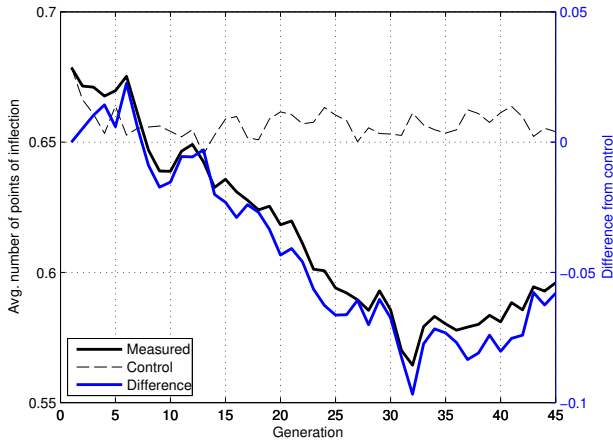
Figure 3. Average number of extrema in the melodic contour.

ENTROPY. It has been hypothesised [10, 11] that simplicity in art is appealing. It hence seems natural to measurewhat happens to the complexity of melodies in the population. To do so, we estimate the average Shannon's entropy [12] of melodies and their rhythms, assuming that intervals between notes and note durations are symbols from a finite alphabet. The entropy is simply

$$H(X) = -\sum_i P(x_i) \log_2 P(x_i), \qquad (1)$$

where X is a string of intervals (or durations), $P(x_i)$ is the probability of occurrence of symbol x_i (estimated from frequency).

Figure 1 shows the average population entropy over time (in this and other figures the bold blue line indicates the difference between the experiment and control). We observe that the melodic entropy is noticeably *decreasing*. This indicates that the respondents tend to select simpler melodies. This is in line with predictions of [10, 11].

Remarkably, the *opposite* result is observed for rhythm: the entropy is noticeably *increasing*, which suggests the preference for more rhythmically varied melodies. (The rhythm entropy in the control experiment is not constant due to a defect in the crossover operator: it does not ensure that the average granularity remains the same. However, the difference from the control shows significant increase of rhythm entropy over time.)

REPETITIVENESS. We also investigated whether the repetitiveness is selected for. To measure repetitiveness, we count the number of adjacent identical intervals (and note durations) and normalise by the length of the melodies. Figure 2 illustrates the trend in thus computed average repetitiveness. Again, we observe increase in melodic and decrease in rhythmic repetitiveness.

We note that higher repetitiveness implies lower entropy, but not the other way round: a repeated alternation between two notes (*e.g.* trill) will have low entropy, but not high repetitiveness.
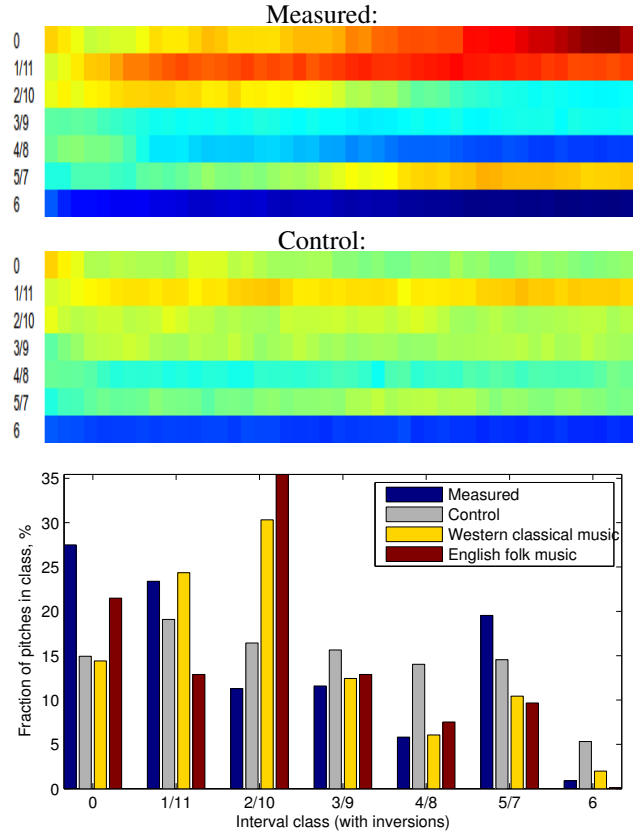

Figure 4. *Above:* Histogram of interval classes (vertically) over time (horizontally); red = high frequency, blue = low frequency . *Below:* the interval class histogram at the last generation.
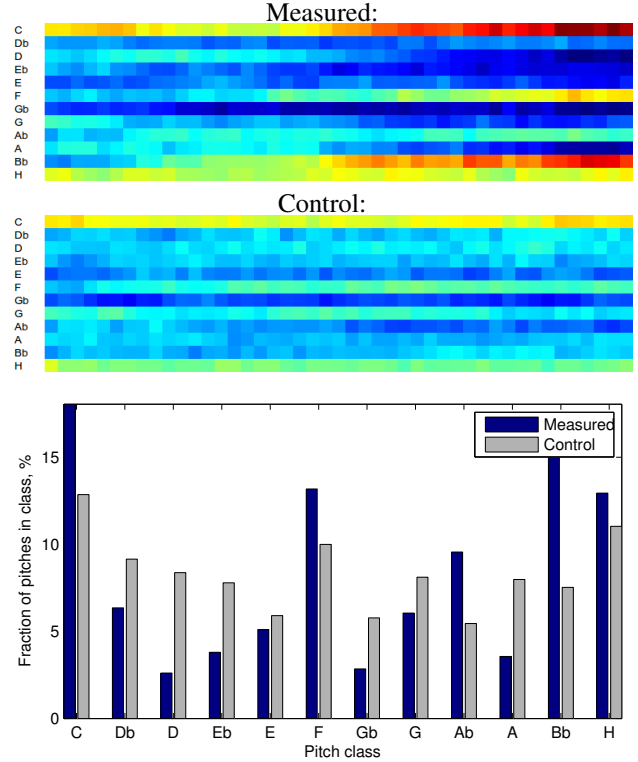

Figure 5. Histogram of pitch classes (vertically) over time (horizontally), red colour indicates higher frequency. *Below:* the pitch class histogram at the last generation.
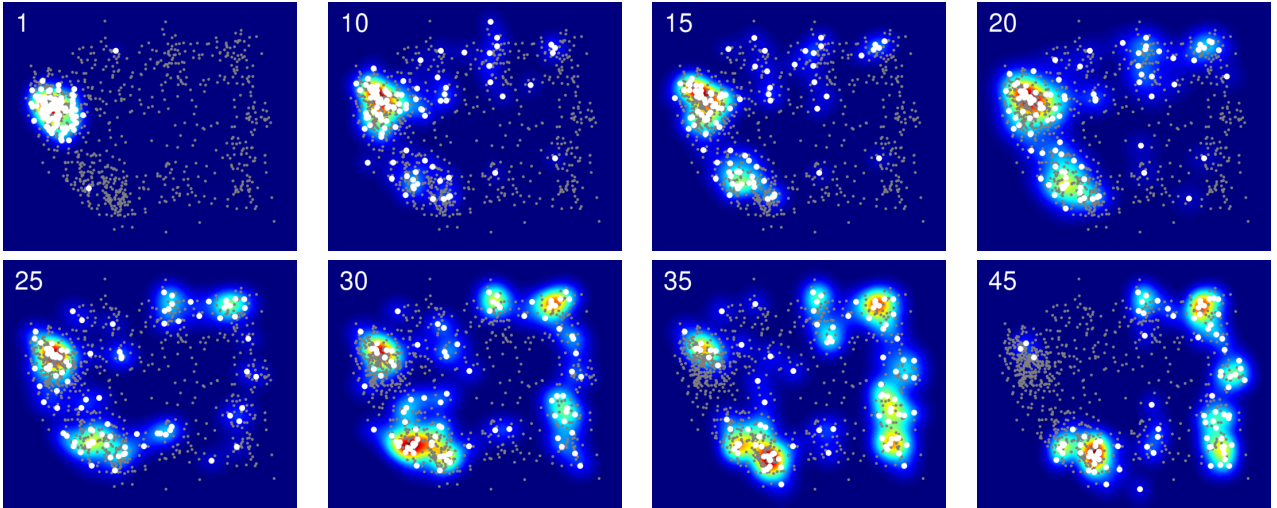
Figure 6. Embedding of the melodies in $\mathbb{R}^2$ over time. The generation numbers are shown in the corner. Pale gray dots represent all melodies ever alive, large white dots — melodies alive in the current generation. Background colour shows the estimated density (using KDE [15]).

MELODIC CONTOUR. To investigate whether the shape of the melodic contour is significant in respondents' selection, we use a simple measure of monotonicity: we compute the number of local extrema in a melody and normalise by the length of the melody. High number of extrema would be indicative of oscillating melodic contours, low number would indicate more monotonously ascending or descending contours. Figure 3 shows the change in average monotonicity over time and illustrates a noticeable decline in this parameter, indicating negative preference for complex, undulating melodies.

PITCH AND INTERVAL DISTRIBUTION. At each generation, we measured the distribution of intervals in the melodies and the resulting pitch classes. Figure 4 (top) shows the evolution of the interval histogram over time. The intervals in Figure 4 are shown modulo 12, and inversions of intervals are placed in the same bin (*e.g.* perfect fourth and perfect fifth, denoted 5/7 in Figure 4, are in the same bin, and similarly for other intervals). We observe a marked preference for consonant, diatonic intervals, and the semitone (0, 1/11, 5/7); the augmented fourth is actively selected agaist (bottom row, 6); interestingly, the prevalence of the major 3rd (4/8) is smaller than that of the major 2nd (2/10) and the minor 3rd (3/9). Figure 4 (bottom) compares the distribution of intervals to that in the control group as well as to that in a corpus of Western classical music [13] and English folk music [14]. We show the correlation matrix for these distributions in Table 1. We observe that the resulting distribution of pitch classes appears to be more characteristic of the Western classical music than of folk music. Figure 5 shows the evolving histogram of pitch classes (top row corresponds to C, second to D-flat *etc.*) Again we observe a marked tendency towards diatonicity: note, for example, the high values in the F-minor triad (F-Ab-C).

EUCLIDEAN EMBEDDING AND CLUSTERING. We regard the space of all melodies as a metric space with the

	Exp.	W.C.	E.F.
Exp.	-	0.0532	0.0497
W.C.	0.0532	-	0.7787
E.F.	0.0497	0.7787	-

Table 1. Correlation between the difference from control in interval class distribution in this experiment vs. that of Western classical music vs. English folk music.

Levenshtein distance [16] (edit distance) applied to strings of intervals forming the melodies (and strings of note durations) as the metric. Having computed pairwise distances between all melodies in the population, we can compute the embedding of the melodies into Euclidean $\mathbb{R}^n$ space. Multidimensional Scaling [17] delivers such embedding, as well as optimal (in the least-squares sense) embeddings into reduced dimensionality spaces $\mathbb{R}^m$, $m < n$. In particular, for visualisation it is convenient to embed the melodies into $\mathbb{R}^2$. Figure 6 shows such embedding. All melodies that ever lived are shown (pale gray), the ones currently alive are marked by large white circles. To illustrate the tendency, in the background we show the distribution density obtained using kernel density estimator [15]. We remark that having started from the initial cluster (Figure 6 top left) the evolution diverges into well-defined directions (downwards and to the right in Figure 6), and new mutually dissimilar stable clusters of melodies are formed. We speculate that this phenomenon is analogous to speciation in biological evolution. Admittedly, a larger scale experiment is required to more accurately study this effect. To illustrate the principal modes in which the evolution progresses, we have clustered all the melodies by similarity (using classic k-means algorithm [18] on the melodies embedded in $\mathbb{R}^n$). Figure 7 show the resulting clusters together with the melodies nearest to the cluster centroids. These correspond to relatively stable melodic "species" in our experiment.
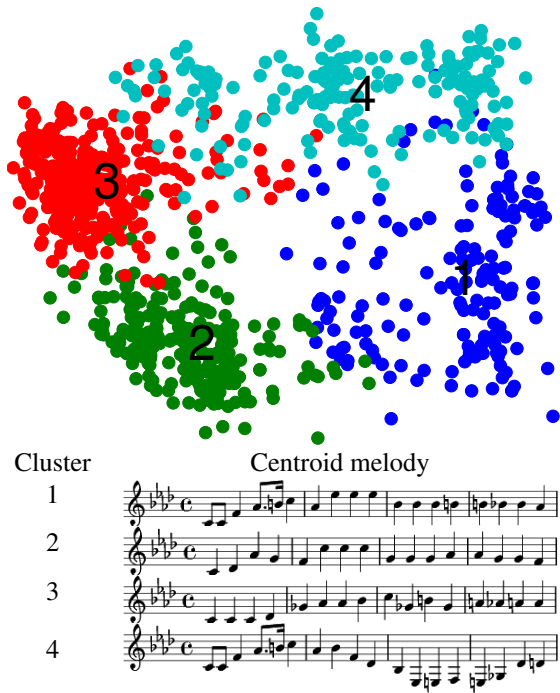
Figure 7. Clusters in the population (*above*) and the melodies nearest to cluster centres (*below*).

Figure 8. Top ranking melody over the generations.

4. DISCUSSION AND CONCLUSIONS

We have conducted an experiment to evolve a population of melodies by popular vote (incidentally, in Figure 8 we show the best ranking melodies over generations). We have performed statistical analysis of the resulting evolutionary history. We have thus used the evolutionary setting in order to examine the popular vote as the fitness function in terms of perceptual biases of the respondents.

By measuring the average entropy, repetitiveness, and variability in melody in rhythm over time, we have come to the conclusion that the respondents are biased towards melodically straightforward, consonant, diatonic, while, on the other hand, rhythmically varied melodies. We have observed that out of chaos emerges the preference for the Western diatonic scale. We also speculate that the formation of stable clusters in diverging branches of evolution may be an effect analogous to speciation. While the above results may already be intuitively familiar to expert musicians, we have for the first time demonstrated that evolu-

tionary setting is a useful tool for studying psychological perceptual biases and æsthetic preferences in humans.

A larger scale experiment would merit more accurate analysis of socio-cultural background of the respondents. It could potentially reveal interesting correlations between the background and musical preferences. It would be also very interesting to conduct a similar experiment with other musical systems, for example those not based on 12-tone scale. Further work would also include a more extensive analysis of the emergent phenomena related to tonality, mode, and key.

Although this experiment is ongoing, and the corpus of data is continuously growing, we believe our preliminary findings may be of interest to the computational music community.

5. REFERENCES

[1] B. E. Johanson and R. Poli, "GP-music: An interactive genetic programming system for music generation with automated fitness raters," U. of Birmingham, Tech. Rep. CSRP-98-13, May 1998.

[2] R. M. MacCallum, M. Mauch, A. Burt, and A. M. Leroi, "Evolution of music by public choice," *PNAS*, vol. 109, no. 30, pp. 12 081–12 086, 2012.

[3] Y. M. A. Khalifa, H. Shi, and G. Abreu, "Evolutionary music composer," in *Late Breaking Papers at the 2004 Genetic and Evolutionary Computation Conference*, M. Keijzer, Ed., Seattle, Washington, USA, 26 Jul. 2004.

[4] J. D. Fernandez and F. J. Vico, "AI methods in algorithmic composition: A comprehensive survey," *J. Artif. Intell. Res. (JAIR)*, vol. 48, pp. 513–582, 2013.

[5] N. Claidière, S. Kirby, and D. Sperber, "Effect of psychological bias separates cultural from biological evolution." *PNAS*, vol. 109, no. 51, p. E3526, 2012.

[6] D. Sperber and L. A. Hirschfeld, "The cognitive foundations of cultural stability and diversity," *Trends in Cognitive Sciences*, vol. 8, no. 1, pp. 40–46, Jan. 2004.

[7] S. Kirby, H. Cornish, and K. Smith, "Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language," *PNAS*, vol. 105, no. 31, pp. 10 681–10 686, August 2008.

[8] K. G. Jamieson and R. D. Nowak, "Active ranking using pairwise comparisons." in *NIPS*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, Eds., 2011, pp. 2240–2248.

[9] F. L. Wauthier, M. I. Jordan, and N. Jojic, "Efficient ranking from pairwise comparisons." in *ICML (3)*, ser. JMLR Proceedings, vol. 28, 2013, pp. 109–117.

[10] J. Schmidhuber, "Low-complexity art," *Leonardo, Journal of the Inter Soc for the Arts, Sciences, and Technology*, vol. 30, no. 2, pp. 97–103, 1997.

[11] N. Hudson, "Musical beauty and information compression: Complex to the ear but simple to the mind?" *BMC Research Notes*, vol. 4, no. 1, pp. 9+, 2011.

[12] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[13] D. Huron, "Tone and voice: A derivation of the rules of voice-leading from perceptual principles," *Musical perception*, vol. 19, no. 1, 2001.

[14] W. J. Dowling, "Rhythmic fission and the perceptual organization of tone sequences," *Musical perception*, vol. 1967 part 1, 1967.

[15] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Annals of Statistics*, 2010.

[16] D. S. Hirschberg, "Serial computations of levenshtein distances," in *Pattern Matching Algorithms*, A. Apostolico and Z. Galil, Eds. Oxford University Press, 1997, pp. 123–141.

[17] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching," *PNAS*, vol. 103, no. 5, pp. 1168–1172, 2006.

[18] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, 1967, pp. 281–297.