LORENZO ROMANELLI

TOWARDS A MODEL OF COMPUTATIONAL BEAUTY

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 STATE OF THE ART

When talking about cognitive sciences, several studies exist that illustrate the underpinnings of human perception and cognition of basic dimensions of sound, such as loudness, pitch, rhythm and timbre (e.g., see Justus and Bharucha, 2002). Further research has focused on higher-level concepts related to music, including the perception of its emotive content and the way in which we tend to express it (Juslin and Laukka, 2004), as well as performance specific traits (Palmer, 1997).

In the Music Information Retrieval (MIR) field, it is useful to categorize these musical dimensions, most commonly referred to as *descriptors*, using a hierarchy organized in three levels of abstraction (Gouyon et al., 2008, among others). Climbing the ladder of such hierarchy up to the top means starting from the most fundamental acoustic features, to be extracted directly from the signal, and progressively building on top of them to get to model more complex concepts derived from music theory, or musicology, or even from cognitive and social phenomena.

The organization of the three levels of this hierarchy, in order of increasing complexity of the features associated with each level, is defined as follows:

1. *low-level descriptors* – loudness, pitch, timbre, onsets, . . .
2. *mid-level descriptors* – tempo, tonality, modality, . . .
3. *high-level descriptors* – genre, mood, instrumentation, . . .

High-level descriptors are referred to also as *semantic* descriptors, for they require an additional induction from users. In other words, we cannot rely solely on data computed directly on an audio signal[1] to define concepts such as the mood of a song. We in fact need to first give an interpretation of terms like *happiness* or *sadness* from the user's perspective, contextualize them, and then approach the study of how these interpretations relate with low or mid-level descriptors extracted from the music. Models used for high-level descriptors have to rely on prior knowledge which is always more or less biased towards the end users of the specific application.

Suppose you had to design an algorithm for a music recommender system based on genre similarity. How would you define which genres are similar to each other? The metrics suited for the task can be many[2]: instrumentation, tempo, rhythm, most likely a mix of them

1 Nor from the symbolic information (usually the score) associated with a piece of music.

2 This is just an example built on common sense; the topic is huge, research on it is abundant and beyond the scope of this work.

and more, or even data which is not necessarily bound to the audio information itself (I am talking about *metadata*). The opinion of a domain expert, even if technically more correct or informed, might be less suited for this application than the perspective of the layman using the platform everyday, who doesn't care if two pieces of music are both from Detroit-based producers.

This problem bound to interpretation of high level semantics is also known as the *semantic gap*. To put it in the words of Smeulders et al., the semantic gap is «[...] The lack of coincidence between the information that one can extract from the (sensory) data and the interpretation that the same data has for a user in a given situation» (Smeulders et al., 2000). The problem of the semantic gap becomes even more relevant as the concept we are trying to model becomes more abstract; this is indeed the context where the problem I am going to outline in the next paragraphs finds its place and some of its practical justifications.

Given the promising advances seen in the field in the last fifteen years, it is surprising to see how the study of the concept of musical beauty from an MIR perspective is barely considered. Dealing with beauty – not only when talking about music – is of course a tricky undertaking. Everyone has heard the common saying that «beauty lies in the eye of the beholder»[3], which, perhaps less poetically, suggests that the experience of the beautiful can only be interpreted as a subjective phenomenon detached from any objective feature of what caused it.

If it was true that we can't explain beauty other than by accepting its independence from any formal, observable and quantifiable property, then transposing the task in an MIR context would have no purpose, and we'd better abandon our hopes. Fortunately, there exists a wealth of research suggesting that a different point of view on the matter might make more sense – not just from a philosophical perspective. Next to aesthetic theories, studies in cognitive neurosciences and artificial intelligence add value to the hypothesis that aesthetic experiences can be explained at least in part by objective foundations.

What follows is a discussion on those pieces of research that I consider relevant for the work of the present thesis, as well as for giving an outline of its limitations; as such, I don't expect it to be taken as an exhaustive literature review on every possible theory about the nature of art and beauty, and even less on the philosophy of aesthetics as a whole[4].

1.1 WHAT IS BEAUTY? A REVIEW OF AESTHETIC THEORIES

Any discourse about beauty must deal with the fact that there isn't a consensus on its nature. This question has been debated for at

3 The quote as we know it appeared for the first time in the book *Molly Bawn* by Margaret Wolfe Hungerford (Hungerford, 1878).

4 For the reader interested in a more comprehensive introduction to Asthetics, I'd suggest to head to other resources; Graham, 2005 and Tatarkiewicz, 2006 are good starting points, which I myself will address multiple times during my discussion.

least 2 500 years and has been given a wide variety of answers. Immanuel Kant thought one premise of beauty was an attitude of "disinterested contemplation" (Kant, 2001 [1790]), whereas Friedrich Nietzsche dismissed this notion and underlined the impact of sensual attraction (Nietzsche, Clark, and Swensen, 1998 [1887])[5]. For the poet John Keats, beauty equaled truth (Keats, 1898), while Stendhal, the French novelist, characterized beauty as the "promise of happiness" (Stendhal, 1927 [1822]). Each of these theories is respected; not one is universally accepted.

In my discussion, I will adopt the *Oxford Dictionaries* definition of beauty as a starting point[6]:

> A combination of qualities, such as shape, colour, or form, that pleases the aesthetic senses, especially the sight.

I don't particularly like this definition, for two reasons. First, there is an explicit reference to its "objective" interpretation, as the term gets bound to concrete qualities and ignores any possible subjective implication. Moreover, this definition suggests that the sight somehow holds a privileged position among the aesthetic senses – whichever those senses are. Does it mean that things that please the eye are to be considered more beautiful than, say, music? Or maybe that we perceive beauty better through sight? Are there more beautiful objects in the visual domain than in others? How can we even quantify beauty[7]?

1.1.1 *Beauty as aesthetic pleasure*

However, not everything should be thrown away. The definition in fact mentions one aspect that is commonly addressed in the philosophical discourse on beauty: beautiful objects cause pleasure to – I would rather say *through* – the aesthetic senses (e.g. Tatarkiewicz, 2006). It is a distinctive kind of pleasure, which exists in a different manner than from the pleasures deriving from a good meal, or fresh air, or a good bath (Ingarden, 1964). For example, the immediate pleasure arising from having a cold drink on a hot day lies exclusively in a positive sensation of the body and has little to do with aesthetic appreciation of the object. In contrast, perceivers look at a painting not to please their body, but to enjoy the painting's beauty (Reber, Schwarz, and Winkielman, 2004). As such, this peculiar type of pleasure is usually referred to as *aesthetic pleasure* (Graham, 2005).

5 There is a whole current of thought, known as *Darwinian aesthetics* or *evolutionary aesthetics*, suggesting that humans may be biologically primed to find particular features more beautiful, because these features may have been selected for optimal survival (e.g., Thornhill, 1998, Grammer et al., 2003), which will not be addressed here.

6 See *Definition of beautiful in English by Oxford Dictionaries*.

7 My interpretation is that, generally speaking, sight is seen – forgive the wordplay – as the most developed of the human senses. Here, this diffused opinion introduces a bias, perhaps to help contextualizing such a broad topic in the limited space allowed by a dictionary.

It has been observed from ancient times that it seems contradictory to describe something as beautiful and deny that we are in any way pleasurably affected by it. As Graham exemplifies, the same thing cannot be said for other concepts such as colours. People usually prefer one colour to another; they can even be said to have a favourite colour, but we could not tell that just by looking at their use of colour words alone. Describing an apple as a «red apple» doesn't imply that I favour red apples over green apples, whereas if I say «a *beautiful* red apple», you immediately get that I am attributing a positive value to that apple[8].

This said, there are important questions arising from the previous observation: can we identify some kind of connection between purely descriptive terms (such as *red* or *green*) and the evaluative term *beautiful*? If so, where does this connection lie? The tradition in Aesthetics tells us that usual answers to these questions fall into one of three currents of thought. I have already hinted at some of them, but let's try to describe the overall picture in a bit more detail.

1.1.2 *Subjectivism, objectivism, interactionism*

The philosopher David Hume is probably the most renowned exponent of the so-called *subjectivist* view, a view which anyways dates back at least to the Sophists (Tatarkiewicz, 2006). It is here that sayings such as «Beauty lies in the eyes of the beholder» and «De gustibus non est disputandum[9]» would find their place. Subjectivists state that beauty is a function of idiosyncratic qualities of the perceiver; which – coming back to the example of colours – is to say that terms like red and green identify real properties of the apple, where instead the term beautiful says something about the person who uses it. This perspective, of course, implies that all efforts to identify the laws of beauty would be futile:

> «To seek the real beauty, or the real deformity, is as fruitless an enquiry, as to seek the real sweet or real bitter.»
>
> (Hume, 1757)

On the opposite, the *objectivist* position sees beauty as a property of an object that produces a pleasurable experience in any suitable perceiver (Tatarkiewicz, 2006). Eduard Hanslick, one of the most respected music critics of the 19th century, states in his foundational book *The Beautiful in Music* that «[...] Although the beautiful exists for the gratification of an observer, it is independent of him (Hanslick, 1957)». This perspective finds one of its earliest theorists as far as Plato; it was incredibly popular in the 16th century, to the extent that artists started introducing books of patterns that other artists could combine with each other in order to create beauty (Gombrich, 1995); and it inspired a great deal of psychological research in the 20th century in the attempt of identifying the critical contributors to beauty (e.g., see Birkhoff, 1933, Arnheim, 1974, Gombrich, 1980, 1995, . . .).

8 And the contrary can be said when using the word *ugly*.
9 Which roughly translates into «Taste cannot be debated».

Between subjectivists and objectivists we can identify a third current of thought, known as *interactionism*. It tends to be the view adopted in most modern philosophical – and not – analyses. What this theory states is that the sense of beauty emerges from patterns in the way people and objects relate (e.g., see Merleau-Ponty, 1964 and Ingarden and McCormick, 1985). Put this way, it is no surprise that interactionism is a favourite among cognitive neuroscientists approaching the study of beauty – this is a relatively young field called *neuroaesthetics* – as it suggests a discrete neural basis (Conway and Rehding, 2013). I will come back to this point later.

Graham, 2005 reports an interesting argument against pure subjectivism, which I will describe in Section 1.4 and to which my research question will be closely related. Graham's point[10] finds its roots in the theory of aesthetic judgments proposed by Immanuel Kant in the *Critique of the Power of Judgment*, first published in 1790. For this reason, in the next section I am going to briefly outline Kant's idea about what kind of judgment is it that results in our saying that something is beautiful.

1.1.3 *Kant's aesthetics*

According to Kant, aesthetic judgments are identified by four distinguishing features. First, they must be *disinterested*: we take pleasure in something because we judge it beautiful, rather than judging it beautiful because we find it pleasurable. The latter type of judgment would be more like a judgment of the *agreeable*, as when we say «I like the taste of avocado».

Aesthetic judgments, in Kant's view, are also both *universal* and *necessary*. This means that the activity of such judgment involves the instrinsic expectation from others to agree with us. We may say that «Beauty is in the eye of the beholder»: but that is not how we act. If I say «I like the taste of avocado», whereas you do not, I can't give you reasons to like the taste of avocado; you just don't. But we do debate about our aesthetic judgements – especially about works of art. What's more, we tend to believe that such debates and arguments can actually achieve something. For many purposes, beauty behaves as if it was a real property of an object, like its weight or chemical composition. But Kant insists that universality and necessity are in fact a product of the human mind[11], in a process that Kant calls *common sense*. The consequence, of course, is that there is no objective property of a thing that makes it beautiful.

Finally, through aesthetic judgments beautiful objects appear to be "purposive without purpose". An object's purpose is the concept according to which it was made, such as a table in the mind of the carpenter. An object is *purposive* if it appears to have such a purpose, or if, in other words, it appears to have been made or designed. It

10 To be fair, his argument seems to be a favourite among those who discard subjectivism, but is not clear who was the first person to bring it forward (probably Thomas Reid, a contemporary of David Hume).

11 This is a similar view to what interactionists propose.

is part of the experience of beautiful objects, Kant argues, that they should affect us as if they had a purpose, although no particular purpose can be found (Kant, 2001 [1790]).

1.2 NEUROAESTHETICS

Recently, in the attempt of understanding even more thoroughly the nature of our appreciation of beauty, a new field of research, known as neuroaesthetics, has started to investigate the correlation between empirical aesthetics and cognitive neuroscience (Pearce et al., 2016). Neuroaestheticians adopt a more grounded approach to the study of beauty than philosophers, in that the former seek to observe recurrent patterns in neurological reactions when the perceiver witnesses acts of beauty. This said, we should not make the mistake of thinking that neuroaesthetics and traditional aesthetics are two completely disjoint fields. I already mentioned in Section 1.1.2 how the interactionist perspective is a favourite among neuroaestheticians (e.g., Juslin, 2013 and Reber, Schwarz, and Winkielman, 2004 are two pieces of research where the authors explicitly take the interactionist side). The influence of Kant's thought appears to be quite dominant as well (Conway and Rehding, 2013).

As it often happens, the first studies in the field have focused on the visual domain. In Kawabata and Zeki, 2004, for example, subjects were shown paintings previously classified by the subjects themselves as "beautiful", as opposed to "neutral" or "ugly". By using a technique known as fMRI (*functional Magnetic Resonance Imaging*), Kawabata and Zeki observed that the perception of different categories of paintings are associated with distinct and specialized visual areas of the brain, that the orbitofrontal cortex is differentially engaged during the perception of beautiful versus ugly stimuli, regardless of the category of painting, and that the perception of stimuli as beautiful or ugly mobilizes the motor cortex differentially.

1.2.1 *Neuroaesthetics of music*

Focusing on music, the work of Brattico and Pearce, 2013 presents a good analysis of the current state of the research. Several neuroimaging studies of musical listening confirm the role of the orbitofrontal cortex in positive affective experiences associated with aesthetic judgments of preference or beauty for music (e.g., see Alluri et al., 2012, Brattico et al., 2011, and Blood and Zatorre, 2001[12]), as it was observed for paintings.

Brattico and Pearce argue (and I tend to agree with that) that there is one important, distinctive difference between neuroaesthetics of art in general (i.e., of visual arts) and neuroasthetics of music, in that the subject of the latter is a complex multidimensional, auditory signal

12 Blood and Zatorre also highlight how pleasure tends to accompain experiences of beauty, providing an empirical motivation to what has been discussed in Section 1.1.1.

extended in time and processed in distinct neural pathways from visual stimuli. One consequence of this distinction lies in the specific focus that must be called for in a neuroaesthetic of music on the role of time: a piece of music cannot be viewed as a static entity, but rather one that unfolds in time, generating and manipulating expectations[13] and interpretations in order to induce an aesthetic experience.

In Brattico and Pearce's conclusions, it is acknowledged the fact that neuroaesthetics of music is still a field in its infancy, and that more empirical research is needed in order to clarify its effectiveness, as well as the practical scenarios where such knowledge could be useful for. They also draw from psychological research to restate the three main factors contributing to an aesthetic experience: the characteristics of the listener, of the listening situation, and, of course, of the music itself. While it is known that all of them assume an important role in defining the aesthetic experience of music (e.g., see Hargreaves and North, 2010), it still is not clear their reciprocal influence, nor in which measure their relative combination contributes to the experience as a whole.

1.3 COMPUTATIONAL BEAUTY

We observed how neuroaesthetics, although with some limitations, can provide us with useful information regarding our neurological reactions when we witness acts of beauty. If it is true that specific brain activity is observed in these situations, not so much we can say about whether these activities are caused by specific properties of the artistic – specifically musical – object. Research in computer science and artificial intelligence (AI) has produced some (more or less valuable) results and theories, in some cases drawing from neuroaesthetics itself.

Once again, the domain of the visual arts has been the one where studies have been the most prolific. In fact, results show how several objective key properties seem to be present in beautiful images. Jacobs et al., 2016 observed that some of these properties correspond to lower spatial frequencies, oblique orientations, higher intensity variation, higher saturation, and overall redness.

Schifanella, Redi, and Aiello, 2015 developed a model which was able to surface beautiful but unpopular pictures from a pool of items uploaded to the photo-sharing platform Flickr. Their approach is based on computing specific descriptors related either to color (e.g., contrast, hue, saturation), spatial arrangement (e.g., symmetry, rule of thirds), or texture (e.g., entropy, energy, homogeneity), and comparing them against the same features computed from a ground-truth of

13 A framework for linking expectations based on statistical learning to aesthetic responses has been proposed in Huron, 2006. According to Huron, an event that is unexpected but ultimately innocuous is capable of inducing a negative prediction response that increases, in a process called *contrastive valence*, the relatively positive limbic effect of the subsequent reaction or appraisal responses. Empirical evidence supports the theory that confirmation or violation of expectations is capable of leading to aesthetic experiences (e.g., Vitz, 1966, Crozier, 1974).

crowdsourced pictures previously labelled as beautiful. As in the case of Kawabata and Zeki, 2004 mentioned in Section 1.2, here the meaning of the term "beautiful" is not defined a priori; it was left to the users' own interpretation. Therefore, by not giving an explicit definition of beauty, we run the risk of including in the aesthetic judgment process a wide variety of criteria (such as preference, stylistic familiarity, popularity, memory, sympathy, elation. . .) whose contribution to aesthetic experiences has not been fully explained yet.

Some theories that try to quantify beauty in music, or at least to give some related measure, have already been proposed. Manaris, Purewal, and McCormick, 2002, and Manaris et al., 2005, for example, conducted experiments exploiting a statistical technique known as Zipf's law on a corpus of MIDI-encoded pieces, suggesting that this technique might be used as a metric for aesthetic evaluation. The music pieces used in their experiments were reportedly selected «by a member [...] with an extensive music theory background», are all pieces belonging to the classical music genre (as much as the vagueness of this label implies), and have been cut down to two minutes, to prevent fatigue in the listeners. I believe that these choices, for which no justification has been provided, could introduce a strong bias to the experiment, since many assumptions are implicitly made here, or not explicitly discarded.

Hudson, 2011 advances an hypothesis that roots in information theory, proposing that compressibility and music appreciation are strictly bound. More specifically, the cognitive process of finding patterns more or less hidden inside a piece of music directly relates to a reward system responsible for our appreciation of it. This hypothesis, although fascinating, lacks the support of empirical experiments, and should therefore be taken with a grain of salt. A related study by McDermott, Schemitsch, and Simoncelli, 2013 shows that the auditory system tends to summarize temporal details of sound textures using time-averaged statistics, especially when the length of the sound is moderate to high.

Brattico, Brattico, and Vuust, 2017, on the same line, and drawing from the studies in visual aesthetics, put forward the hypothesis that our auditory system extracts global features from musical stimuli, and then passes them to the high-level processing responsible for the outcomes of the musical experience, including aesthetic judgment. These global features, analogously to visual features, are defined in terms of distribution of spectral energy, musical texture, expressivity, tempo and mode, and more. Moreover, they propose that the creation of musical beauty is not limited to any particular style, method, genre, or form, implying that the aforementioned model could be applied to any piece of music.

1.4 THE RESEARCH QUESTION

In the previous sections, I have briefly outlined some theories and approaches about beauty and aesthetic judgments. In the discussion

I explained some of the many points of view presented from the perspective of a multitude of disciplines. By now, I hope the reader became aware of how incredibly complex and faceted the topic is, and how anyone willing to tame the problem even from a computational point of view should always at least provide the context they intend to work in, as many variables – such as the methodology or interpretabilty of the results – can be affected by these choices.

The apparent impossibility to find a way out from this labyrinth of opinions, studies, hypotheses should not discourage us to stop investigating; I rather see it as an indicator of the relevance of the problem as well as of the ongoing discussion around it. People, regardless of what sayings tell us, *do* argue over art, over music, over their own preferences, over beauty. Not only that: for the practical purposes of buying paintings and sculptures, judging flower competitions, awarding fashion prizes, granting scholarships, people *need* to argue. We want to award the prize to the most beautiful roses, we want to choose the most beautiful painting submitted in the competition, we want to buy the most beautiful recording of a piece of music, and so on and so forth. There are critics who make a living discussing the relative merits of films, musical compositions, concert performances, paintings, plays and novels. The analysis of *how* people argue over art is a task which I feel deserves more research efforts, especially given the impressive advancements in AI and natural language processing techniques. In the present work, we have to draw some limits: we want to limit the scope of this research to music[14] and, of course, to beauty.

At the end of Section 1.1.2 I hinted at Graham's reasons against subjectivism. He argues the following:

> «[...] In adducing reasons for my preference for a work of art (as for any object over which rational judgement ranges), there is at least one constraint that I am rationally obliged to acknowledge, the need to refer to features that the work actually possesses. I cannot plausibly say that I do not like *The Waste Land* because I do not like limericks, for the obvious reason that *The Waste Land* is not a limerick; I cannot give it as my reason for liking pre-Raphaelite painting that I prefer abstract to representational art, since pre-Raphaelite painting is as far from abstract art as one can get; I cannot justify my distaste for modernist architecture in terms of a more general dislike of excessive or-

14 Someone once said: «Writing about music is like dancing about architecture»; only God knows how much I disagree with that. Robert Christgau gives a nice witty answer to those who so affirm:

> «One of the many foolish things about the fools who compare writing about music to dancing about architecture is that dancing usually is about architecture. When bodies move in relation to a designed space, be it stage or ballroom or living room or gymnasium or agora or Congo Square, they comment on that space whether they mean to or not.»
> (Christgau, 2005)

namentation, because famously modernist architecture eschews ornamentation; and so on.»

(Graham, 2005 – Chapter 11)

What Graham is telling us is that any aesthetic judgment must be carried out according to the actual features of the work about which it is a judgment. Otherwise, we would be talking about matters of mere preference, or personal taste. In other words, expressing an aesthetic judgment (i.e., saying that something is beautiful or ugly) is fundamentally different from statements such as «I like the taste of avocado» – what in Hume's language could be defined as an *original existence*: something that can be acknowledged, but about which not much more can be said. Furthermore, if calling something *beautiful* was equivalent to expressing a simple preference, then why not simply doing so? When I say «This is a *beautiful* piece of music», why would I bother using a term in such a misleading objectified form, as if it was about the piece of music itself, when in fact it is only about me and my feelings towards it?

To wrap up, the points that will be taken for granted from now on, for the reasons discussed in this chapter, are:

1. there is no agreement over the nature of beauty;

2. because of this, it is hard to provide a unique definition of beauty;

3. however, people talk about beauty;

4. when expressing an aesthetic judgment, it is advisable to relate it to real properties of the object of the judgment;

5. the act of giving an aesthetic judgment seems to imply the attribution of both *(a)* a positive or negative value to the object, and *(b)* an objectified status to the judgment itself.

If we hold true these assumptions, and restricting our scope to music, I question whether there exist concepts that people tend to refer to when talking about musical beauty – the "real properties" mentioned in point 4 of the previous list – and, if so, whether it is possible to obtain them in a direct, automatic way starting from unstructured text sources, be they music reviews, comments about songs, playlists descriptions, etc. Thanks to the Internet, there are huge amounts of this kind of data we can take advantage of, while the field of natural language processing (NLP) offers us powerful techniques to extract information from such unstructured data.

I believe that incorporating an analysis of the proposed type into the already existing and ongoing research in philosophy, neurosciences, and computer science can contribute with valuable insights over real case scenarios, insights that would otherwise need to be harvested over more conventional (and with less broad scope, although maybe more controlled) mediums, such as surveys or interviews.

1.4.1 *Limitations*

There are at least two dimensions in aesthetic judgments that have not been mentioned yet whose contribution must be held in mind, which I will here refer to as the *dimensions of variability* of aesthetic experiences.

The first dimension of variability has to do with the observation that the majority of the studies presented here find their context within the boundaries of a Western tradition. The existence of differences between Eastern and Western aesthetics is a generally accepted notion, due to the fact that in non-Western societies aesthetics are more closely related to the communication of spiritual, ethical and philosophical meaning than in the Western tradition (Anderson, 1989).

The second dimension lies in the temporal variable. Aesthetic experience varies throughout historical periods (Pearce et al., 2016), as cultural conventions have shifted or expanded. There are countless examples of artworks which were popular in their day, but whose reputation has since fallen into obscurity, as well as there are examples of artworks which, on the other hand, have caused outrage as soon as they were unveiled in all of their unconventional nature, but have since become admired staples of the repertoire (Igor Stravinsky's *Le Sacre du Printemps* comes off the top of my head).

Therefore, the cultural and historical constitution of the concept of aesthetic experiences should be acknowledged. The choice of our data sources, as we will see in the next chapter, will be subject to these two limitations, as will be the generalizability of the results.

2 METHODOLOGY

The problem described in the previous chapter can be summarized in one sentence:

> *Is it possible to build a model able to capture the topics or concepts commonly addressed when talking about musical beauty?*

As a first step towards finding an answer to this question, we will take advantage of well studied NLP techniques and apply them to a collection of music reviews. The path we will follow for doing so is to obtain a structured representation of the words contained in such reviews, so that mathematical properties of the resulting *semantic spaces* can be exploited to uncover existing semantic relationships between the modeled terms. By querying this model with input words closely related to beauty, we will obtain a set of words which, according to the model's internal representation, are the most semantically related to the input. Finally, we will try to classify the returned similar words, to check whether recurring topics will emerge.

The first phase consisted of gathering the required data. In the following section, I am thus going to describe the adopted dataset.

2.1 A DATASET OF PITCHFORK ALBUM REVIEWS

Pitchfork[1] is a music-centric online magazine, launched in 1995 by Ryan Schreiber and currently based in Chicago, Illinois. It grew out of independent music reviewing into a general publication format. According to the company[2], the website receives «[...] more than 7 million monthly unique visitors».

For our research we will start from a collection of 18 393 Pitchfork album reviews that have been previously scraped from the web and made openly accessible on the Kaggle platform[3]. The collected reviews span an 18-years period, with the earliest having been published on the 5th of January, 1999, and the most recent on the 8th of January, 2017. Reviews were written by 432 different reviewers.

The albums reviewed belong to 8 633 different artists, and each album has been reviewed only once. An album is characterized by zero, one or more genre labels, as summarized in Table 2.1.

Additional pieces of information provided by the dataset include the score given by the reviewer to an album (on a scale from 0 to 10), the record label under which the album has been published, and the content itself of the review, which constitutes the most relevant bit of data for our purposes.

1 https://pitchfork.com/
2 See *Pitchfork | Advertising*
3 https://www.kaggle.com/nolanbconaway/pitchfork-data

GENRE	ALBUMS	GENRE	ALBUMS
rock	9 436	metal	860
electronic	3 874	folk/country	685
experimental	1 815	jazz	435
rap	1 559	global	217
pop/r&b	1 432	*<unlabeled>*	2 367

Table 2.1: Genre labels of the reviewed albums. The *genre* column indicates the label of the genre; the *albums* column indicates the number of albums associated with that label.

2.2 WORD EMBEDDINGS

What we have at disposal is thus an extended collection of documents, in free-text form, from which we wish to extract the closest terms to some input set of words related with beauty. For doing so, we first have to represent the words contained in the documents in a way that allows us to easily compute distances between terms in an unsupervised manner (i.e., without human intervention). The most suitable approaches to achieve this involve using the so-called *word embeddings*.

Under the umbrella name "word embeddings" are included a variety of NLP techniques aimed at mapping words – or in some cases even entire sentences – from a vocabulary onto vectors of real numbers. In mathematical language, we can define a word embedding in the following way:

$$V \to \mathbb{R}^D : w \mapsto \vec{w}$$

meaning a word embedding is a mapping that maps a word w from a vocabulary V to a real-valued vector $\vec{w}$ in an embedding space of dimensionality D. In the simplest case, the vocabulary would be built from the collection of all the single words used in the reviews taken only once.

In order to achieve this task, we have focused our attention on two classes of models. The first one is based on *co-occurrence matrices*, while the second one is known as *word2vec*.

2.2.1 *Co-occurrence matrices*

A co-occurrence matrix is a simple data structure in a matrix form holding how many times any term appears in the same context with every other term in the vocabulary. The assumption is that the more often two terms appear in the same context, the more similar their vector form is, and, consequentially, the more similar they are according to the model.

To compute the co-occurrence matrix, first we need to build a *bag-of-words* (BOW) representation of the words of the vocabulary. A bag-of-words is defined as a matrix $\mathbf{A}$ of size $M \times N$, where M is the number of documents in our collection, and N is the number of terms in our vocabulary. The element $a_{m,n}$ of the matrix holds how many times the term n appears in the document m. The similarity matrix **S** based on term-term co-occurrence is then the result of the multiplication of the document-term matrix **A** by its transposed form:

$$\mathbf{S} = \mathbf{A}\mathbf{A}^{\mathsf{T}}$$

Each row of **S** can be seen as a multidimensional vector representing word n, defined in function of its co-occurrence with all the other words in the vocabulary. As such, we can obtain a single similarity value between any two words by computing the *cosine similarity* of their representative vectors **s** and **t**:

$$\text{similarity} = \frac{\mathbf{s} \cdot \mathbf{t}}{\|\mathbf{s}\|\|\mathbf{t}\|} = \frac{\sum_{i=1}^{n} \mathbf{s}_i \mathbf{t}_i}{\sqrt{\sum_{i=1}^{n} \mathbf{s}_i^2}\sqrt{\sum_{i=1}^{n} \mathbf{t}_i^2}} \tag{2.1}$$

Cosine similarity can vary between 0 and 1, where 0 indicates that the two word vectors are completely dissimilar, and 1 that the two word vectors are the same.

Semantic and syntactic relationships generated in this way can be quite powerful; unfortunately, the drawbacks of applying it on such a big corpus pose serious limits, preventing us from adopting it on the totality of our data. In fact, given the high number of documents and the size of the vocabulary (more than 300 000 unique words), **S** results in an enormous sparse matrix of more than 90 *billion* entries. While preprocessing the text can partially help[4], the amount of information to process is still too demanding in terms of both time and, most importantly, memory.

There exist more advanced techniques that build on top of co-occurrence matrices, such as *latent semantic analysis* and its probabilistic variation (LSA and PLSA respectively, see Hofmann, 1999), but we will not adopt them for the same reasons outlined above.

2.2.2 *Word2vec*

Word2vec is a collection of two related models for computing continuous vector representations of words from very large datasets. They have been presented and further refinished in Mikolov et al., 2013a and b. The architecture of both models consists of a shallow neural network with a single hidden layer. What we are interested in are the weights learned by this hidden layer once the training of the model has been completed.

The difference between the two models of word2vec lies in the way they compute the hidden layer. The first model, called *continuous bag-of-words* (CBOW), aims at predicting a target word by looking at its context words, whereas the second model, called *Skip-gram*, follows

4 During this phase we applied standard stop-words removal and stemming.

the inverse path: given the target word, it will try to predict its context[5].

According to Mikolov et al., CBOW performs better and faster with larger amounts of data; Skip-gram is better suited when the size of the dataset is smaller, and when the amount of rare words is bigger. For these reasons, we chose to adopt the latter model. Even though the amount of words contained in our corpus is notable (more than 12.6 million terms[6]), word2vec is known to produce meaningful results only when the size of corpora is in the order of tens of millions words upwards. In other words, the size of our dataset is barely enough.

Input layer and output layer both consist of W neurons, where W is the number of words in the vocabulary of the given text corpus. The hidden input layer consists of n neurons, where n is another hyperparameter of the model and defines the dimensionality of the vector representation of each word we wish to obtain. Figure 2.1 illustrates a dummy example of a Skip-gram model while it is being trained on predicting the context of the word "ant".

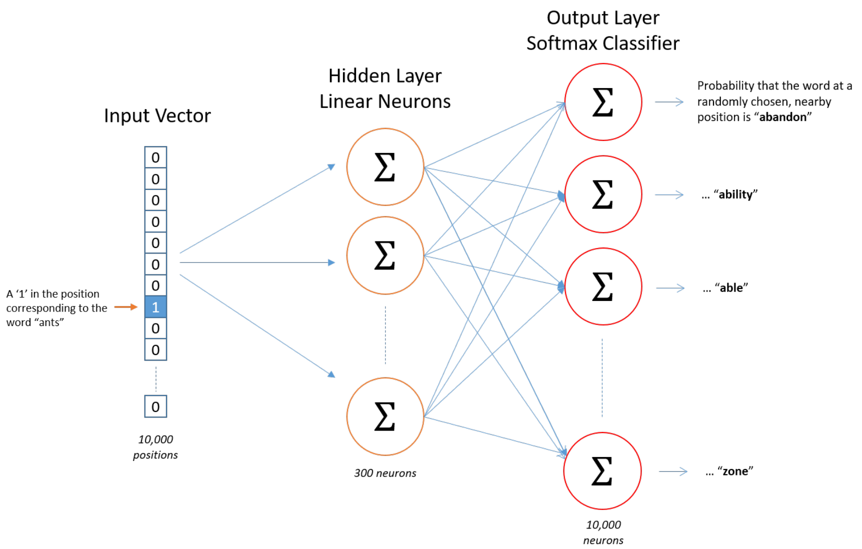

Figure 2.1: Example of a Skip-gram architecture. We can observe that $W = 10\,000$ and $n = 300$, meaning *(a)* that the input vocabulary contains 10 000 terms, and *(b)* that for every word we wish to obtain a 300-dimensional vector representation. Here, the hyperparameter defining the window size is not shown.

The input layer (also called *projection layer*) receives words as a *one hot encoded* vector, i.e. a vector of length v where each element is equal to 0, except for the element whose position corresponds to the position of the input word in the vocabulary. If the vocabulary contains 10 000 words, and the word "ant" appears in it at position 8, the input vector will contain all zeroes, and a single 1 in the 8th element.

---

5 We can define context words as the words to the left and to the right of our target word. A *window size* hyperparameter will tell the model how many context words should be taken into account during the training process.

6 Note that the amount of terms in the corpus and the size of the vocabulary mentioned in Section 2.2.1 are different, since in the vocabulary we only account for unique words.

A more formal definition of the Skip-gram model is as follows. Given a sequence of words $w_1, w_2, w_3, \ldots, w_T$, i.e. the words in our vocabulary, the objective of the model is to maximize the average log probability, defined as:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leqslant j\leqslant c, j\neq 0}\log p(w_{t+j}|w_t) \tag{2.2}$$

where c is the window size. Skip-gram ideally defines $p(w_{t+j}|w_t)$ using a *softmax* function:

$$p(w_O|w_I) = \frac{\exp\left({\mathbf{v}'_{w_O}}^{\mathsf{T}}\mathbf{v}_{w_I}\right)}{\sum_{w=1}^{W}\exp\left({\mathbf{v}'_{w}}^{\mathsf{T}}\mathbf{v}_{w_I}\right)} \tag{2.3}$$

where $\mathbf{v}_w$ and $\mathbf{v}'_w$ are the input and output vector representations of the word w, and W is the number of words in the vocabulary.

However, this formulation is quite expensive, because its computing cost is proportional to the number W of words in the vocabulary (which, in our case, we know to be around 300 000). Mikolov et al., in order to approximate the full softmax[7], propose a more efficient method, the *negative sampling* (NEG). NEG is defined by the following objective function, which will replace every $\log P(w_O|w_I)$ term in Equation 2.2:

$$\log\sigma\left({\mathbf{v}'_{w_O}}^{\mathsf{T}}\mathbf{v}_{w_I}\right) + \sum_{i=1}^{k}\mathbb{E}_{w_i\sim P_n(w)}\left[\log\sigma\left(-{\mathbf{v}'_{w_i}}^{\mathsf{T}}\mathbf{v}_{w_I}\right)\right] \tag{2.4}$$

Here, $\sigma(x) = 1/(1+\exp(-x))$. The task of this optimization is to distinguish the target word w_O from random draws from the noise distribution $P_n(w)$ using logistic regression. The hyperparameter k defines the *negative samples*, or how many words from the noise distribution will be chosen to be "distinguished" from the target word w_O. The noise distribution $P_n(w)$ has been set by the authors as

$$\frac{U(w)^{3/4}}{Z}$$

where $U(w)$ is the *unigram distribution*, a value that in their experiments has been observed to outperform both the unigram and the uniform distribution.

A perhaps more intuitive explanation of the Skip-gram architecture with negative sampling is the following. Whenever the model receives an input word w_I as a one hot encoded vector $\mathbf{h}$, it retrieves from the projection layer the neuron $\mathbf{n}_i$, where i is the index of the only element equal to 1 in $\mathbf{h}$ (the position of the word inside the vocabulary). This neuron holds the weights of the vector representation of w_I. Note that before the training starts, each neuron in the projection layer has to be initialized, usually with small random values.

7 The first version of word2vec uses another approximation of the softmax, the *hierarchical softmax*, not discussed here.

In an ideal scenario (i.e., when using softmax, see Equation 2.3), during training each vector representation of all the words w_O in the vocabulary should either be "pulled closer" to $\mathbf{n}_i$ or "pushed away" from $\mathbf{n}_i$ by a small fraction, depending respectively on whether w_O belongs to the context of w_I or not. With negative sampling, however, the operation of "pushing away" the out-of-context word vectors is not performed for every word in the vocabulary, but only on a small subset randomly chosen from all of the out-of-context words.

There is a further optimization worth mentioning, due to the fact that it will affect data preprocessing. Raw textual sources contain a high number of words which do not carry much information, such as articles and prepositions. For example, if the model will benefit from observing the co-occurrence of the term "guitarist" with the term "guitar", it will benefit much less from observing the co-occurrence of the term "the" with the term "guitarist", since almost every word can co-occur frequently with "the" in a sentence. For this reason, Skip-gram subsamples frequent words according to the following equation:

$$P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}} \tag{2.5}$$

meaning that each word w_i in the training set will be discarded with a probability $P(w_i)$. The term $f(w_i)$ is the frequency of the word w_i, and t is an arbitray threshold, set by the authors at around 10^{-5}.

The advantages of adopting the Skip-gram model are several:

- we can perform a *streamed* type of training, meaning that less computational resources are needed since we won't have to keep all of the corpus loaded in memory all the time; in fact, we can feed the model with one sentence at a time, and then discard it when the next sentence comes in.

- supposing the generated vector representations contain 300 elements each[8], the output matrix computed on our vocabulary will only contain about 90 million entries, corresponding to 0.1% of the size of what we would get by using simple co-occurrence matrices;

- the amount of needed preprocessing is minimal, because, as we have seen, Skip-gram already contemplates a mechanism for discarding irrelevant terms; moreover, processes such as stemming[9] and lemmatization[10] become less relevant, because the model should implicitly figure out that terms sharing the same stem or lemma are in some way similar.

8 This is an indicative value most people tend to suggest as an upper limit, after which overfitting will likely occur, but the choice of the vector size depends on the application.

9 Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form (e.g., the words "beauty", "beautiful" and "beautifully" share the same word stem "beauti").

10 Lemmatization is the process of grouping together the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma, or dictionary form (e.g., the words "play", "plays" and "played" share the same lemma "play").

PARAMETER	DESCRIPTION	VALUE
size	Size of each word's vector representations	300
window	Maximum distance between the current word and the predicted context words within a sentence	5
negative	Number of negative samples	5

Table 2.2: Hyperparameters of the word2vec Skip-gram model

This said, the main reason supporting our choice of relying on this model is that the resulting vector representations will not only generate a semantic space where similar words end up close to each other, but they will be able to represent multiple degrees of similarity between words (Mikolov, Yih, and Zweig, 2013). For our purpose, and given the difficulties encountered in attributing to aesthetic terminology a universal meaning, we could maybe expect to observe one of two scenarios: words such as "beauty" or "beautiful" will *(a)* be very similar to many different (musical) categories of terms (belonging to emotions, instruments, genres, ...), or *(b)* live in a rather isolated corner of the output semantic space, distant from any specific/recognizable category of items.

2.2.2.1 *Preprocessing and training*

Before training the model, it is necessary to build the vocabulary we wish to represent. It has been said before that usually the vocabulary of a dataset is the collection of the single terms used in the documents; however, Mikolov et al. suggest to include in the vocabulary idiomatic phrases whose meaning does not derive from a simple composition of the individual words – what in technical language are referred to as *n-grams*. Examples of music inspired n-grams would be "electric guitar", or "hip hop", or "Guns'n'Roses". Therefore, we first generated a list of n-grams (up to phrases of three words, or trigrams[11]) taken from our corpus that were added to the vocabulary.

Once we defined a vocabulary, we finally trained the Skip-gram model on a lowercased copy of the dataset. Lowercasing raw text is a common preprocessing step in NLP tasks aimed at data cleaning (e.g., words appearing at the beginning of a sentence, or typos). The hyperparameters used in the training are reported in Table 2.2.

2.3 QUERYING AND CLUSTERING

Any word embedding will generate a semantic space, a high-dimensional projection of the vocabulary where every word is represented by vectors. These vectors can be seen as points occupying a specific position inside this multidimensional space. As such, we can apply

11 N-grams can make the model more expressive, but they will also increase data sparsity, so we should be careful and use them with care.

standard clustering techniques to further describe the semantic space, and to characterize the similarities between words. If it is possible to succesfully cluster together words that appear to share a semantic connection, it means that there are good chances the embedding contains organized information, for which the clustering itself can provide some degree of explanation[12].

For this task, what we did was to query the Skip-gram model trained on our data with a list of "aesthetic terms", in order to obtain the closest words to each input query. This simple list comprises the following groups of words:

- aesthetic – aesthetics
- beautiful – beautifully – beauty
- ugliness – ugly

which are very explicit terms related to aesthetics or to beauty (along with their antonyms, adjectives and adverbs).

We finally used a k-means algorithm to cluster the "close terms", or *nearest neighbours*, returned from the model. The nearest neighbours of a word w are all words $v \in V \setminus \{w\}$ (where V is the vocabulary) sorted in descending order by similarity(w, v); the similarity, again, is defined by the cosine distance between vector representations of w and v (see Equation 2.1). By doing so we tried to identify semantic classes of word clusters that could reasonably answer our question: what do people refer to when they argue over musical beauty?

2.4 FINAL NOTE

All the code and the data used in the steps reported here and in the next chapter are open-source and accessible at the following link:

https://github.com/lorenzo-romanelli/compbeauty_code

12 In the next chapter, a less hand-wavy method for evaluating the quality of a word embedding will be introduced, along with results obtained on our dataset.

3 RESULTS

In this chapter, we will mainly present the results obtained on our dataset by following the methodology previously discussed. Before doing so, however, we feel like a brief discussion on quality evaluation of word embeddings is necessary.

3.1 EVALUATING WORD EMBEDDINGS

BIBLIOGRAPHY

Alluri, Vinoo, Petri Toiviainen, Iiro P. Jääskeläinen, Enrico Glerean, Mikko Sams, and Elvira Brattico (2012). "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm." In: *Neuroimage* 59.4, pp. 3677–3689.

Anderson, Richard L (1989). *A Comparative Study of Philosophies of Art*. London: Prentice Hall.

Arnheim, R (1974). *Art and visual perception; The new version (expanded and revised)*. Berkeley: University of California Press. (Originally published 1954).

Birkhoff, George David (1933). *Aesthetic measure*. Vol. 38. Harvard University Press Cambridge.

Blood, Anne J. and Robert J. Zatorre (2001). "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion." In: *Proceedings of the National Academy of Sciences* 98.20, pp. 11818–11823.

Brattico, Elvira and Marcus Pearce (2013). "The neuroaesthetics of music." In: *Psychology of Aesthetics, Creativity, and the Arts* 7.1, p. 48.

Brattico, Elvira, Vinoo Alluri, Brigitte Bogert, Thomas Jacobsen, Nuutti Vartiainen, Sirke Katriina Nieminen, and Mari Tervaniemi (2011). "A functional MRI study of happy and sad emotions in music with and without lyrics." In: *Frontiers in psychology* 2, p. 308.

Brattico, Pauli, Elvira Brattico, and Peter Vuust (2017). "Global sensory qualities and aesthetic experience in music." In: *Frontiers in neuroscience* 11, p. 159.

Christgau, Robert (2005). "Writing about music is writing first." In: *POPULAR MUSIC-CAMBRIDGE-* 24.3, p. 415.

Conway, Bevil R and Alexander Rehding (2013). "Neuroaesthetics and the trouble with beauty." In: *PLoS Biology* 11.3, e1001504.

Crozier, John B (1974). "Verbal and exploratory responses to sound sequences varying in uncertainty level." In: *Studies in the new experimental aesthetics*, pp. 27–90.

Definition of beautiful in English by Oxford Dictionaries. URL: https://en.oxforddictionaries.com/definition/beautiful (visited on 08/17/2018).

Gombrich, Ernst Hans (1980). *The sense of order: A study in the psychology of decorative art*.

Gombrich, Ernst Hans (1995). *The story of art*. Vol. 12. Phaidon London.

Gouyon, Fabien, Perfecto Herrera Boyer, Emilia Gómez Gutiérrez, Pedro Cano, Jordi Bonada, Alex Loscos, Xavier Amatriain, and Xavier Serra (2008). "Content processing of music audio signals." In: *Polotti P, Rocchesso D, editors. Sound to sense, sense to sound: a state of the art in sound and music computing. Berlin: Logos Verlag; 2008.*

Graham, Gordon (2005). *Philosophy of the arts: An introduction to aesthetics*. Routledge.

Grammer, Karl, Bernhard Fink, Anders P Møller, and Randy Thornhill (2003). "Darwinian aesthetics: sexual selection and the biology of beauty." In: *Biological Reviews* 78.3, pp. 385–407.

Hanslick, Eduard (1957). "The Beautiful in Music (1854), ed." In: *M. Weitz, Indianapolis*.

Hargreaves, David J. and Adrian C. North (2010). "21. Experimental aesthetics and liking for music." In:

Hofmann, Thomas (1999). "Probabilistic latent semantic analysis." In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 289–296.

Hudson, Nicholas J. (2011). "Musical beauty and information compression: Complex to the ear but simple to the mind?" In: *BMC research notes* 4.1, p. 9.

Hume, David (1757). "Of the Standard of Taste." In: *Essays Moral, Political, and Literary*. Ed. by David Hume. Libertyclassics (1987), pp. 226–249.

Hungerford, Margaret Wolfe (1878). *Molly Bawn*. Coll. of British authors. Tauchnitz ed. vol. 1788-89 v. 1. Tauchnitz.

Huron, David Brian (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Ingarden, Roman (1964). "Artistic and aesthetic values." In: *The British Journal of Aesthetics* 4.3, pp. 198–213.

Ingarden, Roman and Peter J. McCormick (1985). *Selected papers in aesthetics*. Catholic University of America Press.

Jacobs, Richard H. A. H., Koen V. Haak, Stefan Thumfart, Remco Renken, Brian Henson, and Frans W. Cornelissen (2016). "Aesthetics by numbers: links between perceived texture qualities and computed visual texture properties." In: *Frontiers in human neuroscience* 10, p. 343.

Juslin, Patrik N (2013). "From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions." In: *Physics of life reviews* 10.3, pp. 235–266.

Juslin, Patrik N. and Petri Laukka (2004). "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening." In: *Journal of New Music Research* 33.3, pp. 217–238.

Justus, Timothy C. and Jamshed J. Bharucha (2002). "Music perception and cognition." In: *Stevens' Handbook of Experimental Psychology*.

Kant, Immanuel (2001 [1790]). *Critique of the Power of Judgment*. Cambridge University Press.

Kawabata, Hideaki and Semir Zeki (2004). "Neural correlates of beauty." In: *Journal of neurophysiology* 91.4, pp. 1699–1705.

Keats, John (1898). *Ode on a Grecian Urn: And Other Poems*. Houghton, Mifflin.

Manaris, Bill, Tarsem Purewal, and Charles McCormick (2002). "Progress towards recognizing and classifying beautiful music with computers-

MIDI-encoded music and the Zipf-Mandelbrot law." In: *SoutheastCon, 2002. Proceedings IEEE*. IEEE, pp. 52–57.

Manaris, Bill, Juan Romero, Penousal Machado, Dwight Krehbiel, Timothy Hirzel, Walter Pharr, and Robert B. Davis (2005). "Zipf's law, music classification, and aesthetics." In: *Computer Music Journal* 29.1, pp. 55–69.

McDermott, Josh H., Michael Schemitsch, and Eero P. Simoncelli (2013). "Summary statistics in auditory perception." In: *Nature neuroscience* 16.4, p. 493.

Merleau-Ponty, Maurice (1964). *The primacy of perception: And other essays on phenomenological psychology, the philosophy of art, history, and politics*. Northwestern University Press.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (2013). "Linguistic regularities in continuous space word representations." In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–751.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean (2013a). "Distributed representations of words and phrases and their compositionality." In: *Advances in neural information processing systems*, pp. 3111–3119.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013b). "Efficient estimation of word representations in vector space." In: *arXiv preprint arXiv:1301.3781*.

Nietzsche, Friedrich, Maudemarie Clark, and Alan J Swensen (1998 [1887]). *On the genealogy of morality*. Hackett Publishing.

Palmer, Caroline (1997). "Music performance." In: *Annual review of psychology* 48.1, pp. 115–138.

Pearce, Marcus T., Dahlia W. Zaidel, Oshin Vartanian, Martin Skov, Helmut Leder, Anjan Chatterjee, and Marcos Nadal (2016). "Neuroaesthetics: The cognitive neuroscience of aesthetic experience." In: *Perspectives on Psychological Science* 11.2, pp. 265–279.

Pitchfork | Advertising. URL: https://pitchfork.com/ad/ (visited on 08/28/2018).

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman (2004). "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?" In: *Personality and social psychology review* 8.4, pp. 364–382.

Schifanella, Rossano, Miriam Redi, and Luca Maria Aiello (2015). "An Image Is Worth More than a Thousand Favorites: Surfacing the Hidden Beauty of Flickr Pictures." In: *ICWSM*, pp. 397–406.

Smeulders, Arnold WM, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain (2000). "Content-based image retrieval at the end of the early years." In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12, pp. 1349–1380.

Stendhal (1927 [1822]). *On Love*. Trans. by H.B. V. Works of Stendhal. Boni & Liveright.

Tatarkiewicz, Wladyslaw (2006). *History of Aesthetics: Edited by J. Harrell, C. Barrett and D. Petsch*. A&C Black.

Thornhill, Randy (1998). *Darwinian aesthetics.* Lawrence Erlbaum Associates Publishers.

Vitz, Paul C (1966). "Affect as a function of stimulus variation." In: *Journal of Experimental Psychology* 71.1, p. 74.