Efficient Communication in Written and Performed Music

Laurent Bonnasse-Gahot

Efficient communication in written and performed music

Laurent Bonnasse-Gahot

Centre d'Analyse et de Mathématique Sociales, CNRS, EHESS, PSL University, Paris, France

Abstract

Since its inception, Shannon's information theory has attracted interest for the study of language and music. Recently, a wide range of converging studies have shown how efficient communication pervades language, from phonetics to syntax. Efficient principles imply that more resources should be assigned to highly informative items. For instance, average information content was shown to be a better predictor of word length than frequency, revisiting one of the famous Zipf's law. However, in spite of the success of the efficient communication framework in the study of language and speech, very little work has investigated its relevance in the analysis of music. Here, we examine the organization of harmonic information in two large corpora of Western music, one made of MIDI files directly sequenced from scores, and the other made of MIDI recordings of live performances of highly skilled piano players. We show that there is a clear positive relationship between (contextual) information content of harmonic sequences and two essential musical properties, namely duration and loudness: the more unexpected a harmonic event is, the longer and the louder it is.

Keywords: information theory; efficient communication; music; harmony

1 Introduction

In 1948, Claude E. Shannon published his seminal paper *A Mathematical Theory of Communication*, founding the field of information theory, which has had a tremendous impact on many hard science disciplines such as signal processing and computer science (Mac Kay, 2003; Cover and Thomas, 2006), but has also attracted interests and applications, since its very beginnings, in the study of language (Shannon, 1948, 1951; Cherry et al., 1953; Mandelbrot, 1954) and music (Meyer, 1956; Pinkerton, 1956; Meyer, 1957; Brooks et al., 1957; Youngblood, 1958; Cohen, 1962).

A fundamental theorem brought by Shannon states that a good communication code, realized as a sequence of symbols, should match the length of each symbol with its probability. This principle is widely used nowadays in data-compression techniques (eg: Huffman, 1952). But this idea was already intuited before the advent of information theory. For instance, in Morse code, designed by Samuel Morse and Alfred Vail in the mid-nineteenth century, the length of the sequence of symbols used to code a given character is roughly in inverse relationship with the frequency of occurrence of that character in English texts. The most frequent letter, `e`, is thus coded by a single *dot* ($\cdot$), which is the simplest code in Morse, whereas the letter `z`, a much rarer letter, is coded as the longer *dash dash dot dot* ($--\cdot\cdot$). This allows for a more efficient use of the bandwidth, sending more information per unit of time, while also limiting repetitive strain injury to the operator for the same amount of letters transmitted. The same principle

is at work for languages, as linguist George K. Zipf showed about eighty years ago: "the larger a word is in length, the less likely it is to be used" (Zipf, 1936). Zipf does not just exhibit the relationship, he also argues that it is the frequency that shapes the length and not the contrary.

One of the great success of Shannon theory was to precisely and operationally define the common but at the time vague notion of information. Given context c, the information content (also called surprisal) of a particular outcome x of a random variable is:

$$-\log_2 p(x|c) \tag{1}$$

If the logarithm is taken in base 2 (as is the case throughout this work), the unit of information is called the *bit*. First, with this definition, Eq. 1, one immediately sees that the less probable (thus the more unexpected) an event is, the more informative it is. Second, the formula takes advantage of the property that the logarithm of a product is equal to the sum of the logarithms of its factors. This form thus allows to simply write the total information conveyed by two independent events as the sum of the information of each corresponding event. Finally, the notion of context is crucial here. Surprisal of an event depends on the conditions in which this event happens. For instance, as Markov (1913) already noted a century ago, a given letter does not have the same probability of appearance depending on the letters that precede it. Consider for example the letter `h` in English: you will be more surprised to find it after the letter `b` than if it follows the letter `t`.

The past two decades have seen many works looking at language and speech from the point of view of information theory. For instance, the latter statement about word length and frequency was recently refined: the length of a given word is better explained by its information content than by its overall frequency (Piantadosi et al., 2009, 2011) (see also Manin, 2006: the unpredictability of a word is linearly related to its length). Many other aspects that fall in line with this idea have been studied. At the word level, when an instance of a word is informative (*ie* less predictable), speakers tend to increase its duration (Aylett and Turk, 2004, 2006), whereas predictive contexts lead speakers to choose shorter variants of words (e.g. exam vs examination) (Mahowald et al., 2013). Similarly, speakers prefer to choose contracted forms over full forms ("I'm" vs "I am") in more predictive contexts (Frank and Jaeger, 2008). At the prosodic level, stressed syllables are more informative than unstressed ones (Piantadosi et al., 2009), and more informative syllables are produced with longer duration (Aylett and Turk, 2004). At the phonetic level, informative context increases phonetic lengthening, whereas predictive context increases the probability of phonetic reduction and deletion (Aylett and Turk, 2004, 2006; Cohen Priva and Jurafsky, 2008; Cohen Priva, 2015). Finally, predictability was shown to affect discourse at the syntactic level (Levy and Jaeger, 2007; Jaeger, 2010; Temperley and Gildea, 2015). All these results form a diverse and rich set of evidence that supports the uniform information density hypothesis (Fenk and Fenk, 1980; Fenk-Oczlon, 2001; Genzel and Charniak, 2002; Aylett and Turk, 2004; Levy and Jaeger, 2007; Jaeger, 2010), according to which speakers manage the rate of information so as to keep it as even as possible, "avoiding peaks and troughs in information density".

Surprisingly, despite the abundant evidence discussed above in the case of language and speech, very few studies have been conducted in the field of music cognition. To our knowledge, there are only two exceptions. First, the experimental work by Bartlette (2007, as cited in Temperley, 2014) showed that musicians tend to perform unexpected harmonic events with a longer duration than more expected ones. Second, analyzing a corpus of common-practice themes, Temperley (2014) found that in the second instance of a repeated pattern that contains one changed interval, this interval tends be larger and/or the

new pattern tends to be more chromatic. This is interpreted as a way to balance the information flow: the repetition induces low-information content, but larger or more chromatic intervals introduce more surprisal. Yet, despite such scarcity of studies, one could think of several reasons why we might expect to find efficient communication in music too. First, some authors have proposed music as a protolanguage (Darwin, 1871, and see Fitch, 2006, for a discussion), and neural bases of language and music possibly overlap substantially (Patel, 2003; Koelsch, 2005; Fedorenko et al., 2009; Peretz et al., 2015, but see Fedorenko et al., 2011, 2012). Moreover, speech and music share a similar acoustic code in the way it puts emphasis on important events: notably, of special interest for the present work, accents can be realized by an increase in loudness and/or duration (Carlson et al., 1989). Finally, predictability, which lies at the core of the concept of information, plays a major role in music processing (Meyer, 1957; Huron, 2006; Tillmann, 2012; Rohrmeier and Koelsch, 2012; Pearce and Wiggins, 2012).

The current work aims at filling in part this gap by looking at efficient communication in music. Music is a highly complex phenomenon, and many aspects, notably melodic, rhythmic, or harmonic, could be investigated. Here, we focus on harmony, which might be defined as "the study of simultaneous sounds (chords) and of how they may be joined" (Schoenberg, 1978). Harmonic content is here defined in a precise and workable way. Tones one or several octave(s) apart share a similar perceptual savor, called chroma (see Deutsch, 1980, for a review). We consider a piece of music as a sequence of binary chromagrams, each chroma vector being defined as a 12-dimensional vector of pitch classes, with 1 if a note with the corresponding pitch class is present, 0 otherwise. For instance, using a vectorial notation that starts with pitch C and goes chromatically up to B, the most common chromagram in our main corpus is the vector `(100010010000)`, which corresponds to a C major triad (composed of the C, E, and G pitch classes, independently of their specific vertical arrangement, or voicing). This way of looking at harmonic content is somewhat agnostic to harmonic theories, making it possible to notate any chord, notably passing chords or even ones that would be difficult to label within traditional analysis (consider for instance the texture from the Augurs of Spring in Stravinsky's The Rite of Spring).

We restrict ourselves to the study of Western music, mainly within common practice period (baroque, classical, romantic), although we also consider a few composers from early post-common practice. The materials that we exploit consist in a large number of files encoded in the standard Musical Instrument Digital Interface (MIDI) format, freely available on the Web. We consider two corpora:
(i) One that is as close as possible to written music, with MIDI files directly sequenced from the score. In a traditional musical score, many musical indications are indicated in a symbolic form. For instance, dynamics can be marked as pianissimo (pp), mezzo-forte (mf) or forte (f), but the actual realization of this indication is left to the performer. Here, the necessary conversion of a performance marking into a numeric value is either left to the software used to convert the music into the MIDI format, or to the person that manually sequenced the score. In a sense, a musical piece in the MIDI file format is already a performance (albeit very crude) of the piece. Still, the expressiveness of such rendition is rather low compared to a real live performance. As a limitation, note that the translation process of the scores might be subject to irrelevant biases either hard coded into the software or introduced by the transcriber, by adding extra (unwritten) expressiveness or due to the conversion of the existing marking.
(ii) One that consists in MIDI recordings of live performances of highly skilled piano players. These recordings capture the nuances and subtleties in timing, phrasing, and dynamics of these expressive performances.

As detailed below, after segmenting each piece of music into a sequence of harmonic vectors, we evaluate the information content of each such vector, looking at several millions of them, and relate

this surprisal quantity with other properties of the corresponding musical segment, namely duration and loudness. The information content of a given harmonic segment is evaluated as the amount of surprise experienced by a model trained on the same kind of musical material. Following efficient principles, we expect informative items to have more prominence: the higher the information content of a harmonic event, the longer its duration and the greater its loudness.

2 Materials and Methods

2.1 General pipeline

Our workflow is the following. (i) A score or a performance is first turned into a MIDI file. This step is already done here, as we directly work with MIDI files of either sequenced music or recorded live performances, freely available on the Web. (ii) We segment this MIDI representation into harmonic vectors that capture the vertical relationships of the tones, in terms of the pitch classes that are present. Each such vector is also characterized by its duration and its loudness. (iii) We evaluate the information content of each vector as the amount of surprise experienced by a recurrent neural network model trained on the same kind of musical material. (iv) We compare this quantity with duration and loudness.

We share the preprocessed data, as well as the full Python 3 source code used for preprocessing the MIDI files, computing the information content, performing the analyses and visualizing the results. The custom code written for the present project makes use of the following libraries: `keras v2.2.4` (Chollet et al., 2015) (using `tensorflow_gpu v1.14.0`, Abadi et al., 2015), `matplotlib v3.0.3` (Hunter, 2007), `numpy v1.16.2` (Van Der Walt et al., 2011), `pandas v0.24.2` (McKinney et al., 2010), `pretty_midi v0.2.8` (Raffel and Ellis, 2014), `scikit_learn v0.21.3` (Pedregosa et al., 2011), `seaborn v0.9.0`, `statsmodels v0.9.0`. Data and code available on the Open Science Framework at `https://osf.io/gxw4b` (Bonnasse-Gahot, 2019).

2.2 Musical material

Two large corpora are considered, one with sequenced MIDI files and the other with MIDI recordings of live performances. Among the many parameters that make up a MIDI file[1], we are here interested in three quantities: the timing of each note, so as to extract the duration of each segment; the pitch of those notes, whose chroma is turned into the harmonic vector; and finally the velocity, a quantity related to the speed of the hammer that strikes the strings, proportional to the loudness (Jeong and Nam, 2017). The latter parameter, the velocity, is much more meaningful in the case of the performed corpus, not only because each note has its own expressive velocity, but also because this corpus is homogeneous in terms of instrumentation (music written for the piano) and interpretation (music performed on the same instrument, a Yamaha Disklavier Pro piano).

2.2.1 Written music

A sizable corpus of files was obtained from the `kunstderfuge.com` website[2], which hosts a large collection of classical music files in MIDI format. The files were sequenced by different authors from the written score into the MIDI format. We consider 23 composers, from Buxtehude to Stravinsky, spanning a wide musical period, from the 17th to the early 20th century. Each file might correspond to a small piece of music or to a full symphony. All in all, 2066 files were included in the present study, amounting to about

[1] `https://www.midi.org/`
[2] `http://www.kunstderfuge.com/`

composer	birth	death	# files	# chroma vectors
Albéniz, Isaac	1860	1909	61	72295
Bach, Johann Sebastian	1685	1750	157	163760
Beethoven, Ludwig van	1770	1827	102	290562
Brahms, Johannes	1833	1897	116	246243
Bruckner, Anton	1824	1896	30	63991
Buxtehude, Dieterich	1637	1707	82	85915
Chopin, Frédéric	1810	1849	65	88925
Couperin, François	1668	1733	94	57257
Debussy, Claude	1862	1918	93	119804
Dvořák, Antonín	1841	1904	141	277813
Fauré, Gabriel	1845	1924	82	84703
Haydn, Joseph	1732	1809	203	306240
Janáček, Leoš	1854	1928	22	25501
Liszt, Franz	1811	1886	106	165747
Mahler, Gustav	1860	1911	28	108009
Mozart, Wolfgang Amadeus	1756	1791	49	193620
Saint-Saëns, Camille	1835	1921	83	120337
Schubert, Franz	1797	1828	129	273385
Schumann, Robert	1810	1856	85	163163
Scriabin, Alexander	1872	1915	78	45788
Stravinsky, Igor	1882	1971	25	36035
Tchaikovsky, Pyotr Ilyich	1840	1893	195	268714
Vivaldi, Antonio	1678	1741	40	34329

Table 1: List of composers included in the written corpus.

190 hours of music in total, from which we extracted more than 3 millions harmonic vectors. Table 1 presents the list of composers that were considered in this study, along with their dates of birth and death, the number of MIDI files included, and the number of harmonic vectors extracted. All the details about the exact pieces that were considered are available in the preprocessed data supplied with the present work.

2.2.2 Live performed music

All the entries from the 2018 piano-e-competition[3] were included. These files are recordings in the MIDI format of live performances of highly skilled musicians on Yamaha Disklavier Pro pianos. These instruments are able to capture all the minute details of the musical executions, notably the expressiveness of the timing and dynamics, and record them in the MIDI format. Here again, music repertoire is drawn from common practice period. In total, 252 files were included in the present study, amounting to more than 30 hours of music, from which we extracted more than half a million harmonic vectors.

2.3 Data preprocessing

Fig. 1 presents the workflow for the data preprocessing. First, a music score (or a live performance) is turned into a MIDI file representation, where each note has a well-determined numeric value for its onset, its duration and its intensity. This part is already done as we work directly from the MIDI files. From these materials, we then extract the sequence of harmonic vectors. In this step, we first perform a musical segmentation, using a procedure similar to Dubnov et al. (2003) or White and Quinn (2016). More precisely here, we segment the musical flow at each new onset, provided this onset is not (quasi-)simultaneous with the previous one. During a performance, whether intentional (for expressive purposes) or not, the notes constitutive of a chord are never struck at the exact same time, even if the tones are perceived as simultaneous. Onsets within a time window of 50 ms were then considered

[3] `http://www.piano-e-competition.com/ecompetition/midi_2018.asp`

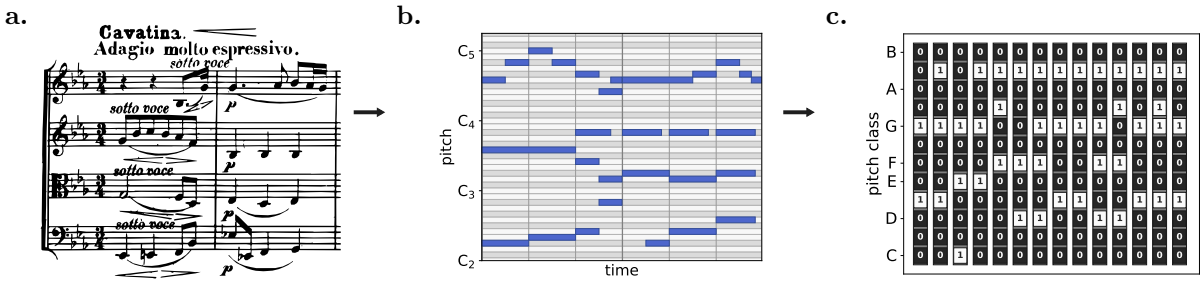

Figure 1: **Pipeline for the preprocessing of the musical data.** Each musical piece, either from a score or from a piano performance, (a), is turned into a pianoroll, (b), in the MIDI format, which is finally segmented and transformed into a sequence of pitch class vector, (c).

synchronous and were fused, so that only one harmonic vector got associated with that group of onsets. A segment is then defined as the musical content comprised between two different onsets. The harmonic vector associated with this segment is taken as the binary chroma vector, that represents the pitch classes that are sounding during this period of time (see Fig. 1c). Note that rests, defined as periods of silence longer than 1 s, are also segmented, and associated with a chroma vector of zeros. Each segment is also annotated with its duration, thus defined as the time interval between two onsets, and its velocity (proportional to the loudness), taken as the maximum MIDI velocity of the notes being struck at the beginning of the onset of that segment.

2.4 Evaluation of information content

2.4.1 Model

A key step of our study lies in the estimation of the information content of each harmonic segment. We evaluate the information of a chroma vector as the prediction error of a recurrent neural network (RNN) that has been trained on a corpus of the same style. For that, we use the long short-term memory (LSTM) architecture (Hochreiter and Schmidhuber, 1997). These networks have distinctive properties that make them appealing for the present work. First, LSTMs exhibit state-of-the-art performance on many tasks involving time series or sequential data, notably in the case of language modeling (Melis et al., 2018; Merity et al., 2017). Various studies have shown how these neural networks – unlike n-grams – are able to capture humanlike traits in various linguistic settings, such as word order preferences (Futrell and Levy, 2019) or number agreement dependencies (Linzen et al., 2016). In the task of predicting the next time step given the previous ones in a musical sequence, RNNs and in particular LSTMs are known to perform better than other modeling techniques such as n-grams or hidden Markov models (Boulanger-lewandowski et al., 2012; Walder, 2016). Moreover, contrary to traditional RNNs (Elman, 1990), LSTMs are able to learn simple context-free and context-sensitive languages (Gers and Schmidhuber, 2001). Finally, they can better capture long-term dependencies (Eck and Schmidhuber, 2002), which is particularly relevant in the context of music (Lerdahl and Jackendoff, 1985). Knowing that the mean duration of a segment is about 200ms, using n-grams with a workable value of n would amount to consider only a very short temporal context, whereas harmonic understanding requires longer time span (see e.g. Bigand et al., 1999; Koelsch et al., 2013).

We thus consider a LSTM with the following architecture: from the input (the current harmonic vector) and the past states contained in the memory cells of the neural network, the output of the

model tries to predict the next harmonic vector, represented as a 12-dimensional vector that evaluates the probability of each individual pitch class[4]. Following a traditional cross-validation approach, each dataset is repeatedly split into a training set, used during the learning phase of the model, and a test set, left for the evaluation (cross validation is 8-fold here). The learning phase somehow mimics the musical enculturation following experience within a particular style, defined by its own statistics. In order to take advantage of the fact that harmonic knowledge is transposition invariant, we augment each training set by transposing the original content in all twelve tones, which allows the generation of more training data for the model. The information content of each harmonic vector is then evaluated in the test part. We evaluate the information content for each composer separately in the case of the sequenced corpus, and for the whole dataset in the case of the performed corpus.

The error function used during the training and test phase is the binary cross entropy. Assuming conditional independence between each dimension, this error directly corresponds to the information content as defined in Eq. 1. Indeed, first consider one pitch class $x_i \in \{0, 1\}$. Given a certain musical context c, the model evaluates the presence of this chroma as a probability p_i. The associated surprisal $-\log_2 p(x_i|c)$ is thus equal to $-\log_2(p_i)$ if $x_i = 1$ or $-\log_2(1-p_i)$ if $x_i = 0$, which can be more compactly written as:

$$-\log_2 p(x_i|c) = -x_i \log_2(p_i) - (1 - x_i) \log_2(1 - p_i) \tag{2}$$

We can then write the full information content $-\log p(x|c)$ as:

$$-\log_2 p(x|c) = -\log_2 \prod_{i=1}^{12} p(x_i|c) = -\sum_{i=1}^{12} \log_2 p(x_i|c) \tag{3}$$

$$= -\sum_{i=1}^{12} x_i \log_2(p_i) + (1 - x_i) \log_2(1 - p_i) \tag{4}$$

which is precisely the binary cross entropy loss function.

In other words, given a certain context, the information content of a new harmonic vector is equal to the error made in predicting this chromagram. Before learning, surprisal is equal to 12 bits, and there is no relationship between duration and information or between velocity and information.

2.4.2 Technical details

We use a LSTM network with one layer of 128 neurons, with default initial random initialization. Training is performed through gradient descent using Adam optimizer with default parameters (Kingma and Ba, 2015), truncated backpropagation with 32 time steps, with a 0.2 dropout (Srivastava et al., 2014) and a 0.2 recurrent dropout (Gal and Ghahramani, 2016). The cross validation is 8-fold. For each fold, training is done over 15 epochs. We run this simulation 5 times, each time with a different random seed, considering in the end for each chroma vector the mean information value, in order to accommodate

[4] Another way would be to predict each possible chromagram as a separate 'word', as done in language modeling. There are several problems with this method. First, most of the chromagrams are actually rare. If we look at each composer in the written corpus, on average, about 85% of the chromagrams have a frequency inferior to 1‰, and about 25% of them happen only once or twice in each set. Many chroma vectors would then appear as 'out of vocabulary', meaning that they would appear in the test set but not in the training set. This is problematic as it would impair the evaluation of the information content of such rare events, which are not only very common but are also the ones that particularly interest us here. Second, it is well known that chords lie in a lower dimensional space (Krumhansl and Kessler, 1982; Lerdahl, 2004), and this has an impact on their predictability. Consider for instance chord substitution, or chord enrichment, based on common tones. Imagine you expect a CMaj chord and observe a CMaj7, that you might have never seen before: you might be surprised, but not as surprised as if it was a completely unrelated chord. Evaluating the probability of a given chroma vector through the probability of its individual pitch classes makes it thus possible to better approximate the information content of rare or even unseen chromagrams.

for the various sources of randomness (randomness in the initialization of the parameters of the neural network, in the dropout noise, in the stochastic gradient descent and the order of presentation of the training input sequences, in the split of the data into folds for the cross validation).

3 Results

For each corpus, we compute the Spearman rank correlation between information content and duration, and between information content and velocity (loudness). We also look at the relationship between information content and relative duration, defined as the ratio between the duration of the corresponding segment and the average duration of segments within a centered window of 10 s. Likewise, we look at the relationship between information content and relative velocity, defined as the ratio between the velocity of current segment and the average velocity of the 5 center-surrounding segments.

3.1 Written corpus

Fig. 2 presents the results for the written (sequenced) corpus, where all composers are pooled together (see Supplementary Figs. S1 to S4 for details per individual composer). One could first notice that the information content of each harmonic segment is much lower than before training (with a mean value of 5.6 bits compared to the initial value of 12 bits), which indicates that the model has indeed learned something.

We found a significant positive correlation between information and (i) duration ($r = 0.22$, $p <$ 1e−100, $z = 399$), (ii) relative duration ($r = 0.10$, $p <$ 1e−100, $z = 178$), (iii) velocity ($r = 0.07$, $p <$ 1e−100, $z = 124$), and (iv) relative velocity ($r = 0.18$, $p <$ 1e−100, $z = 323$). Two-level analyses that compute statistics for each composer and examine the distribution of the resultant statistic confirm the latter pooled results. For each composer, we computed the (Spearman) correlation between information content and each variable of interest (duration, relative duration, velocity, relative velocity). Exact values of these correlations are provided in Supplementary Figs. S1 to S4 (see also Supplementary Fig. S5 for a visual summary). For each variable, the 95% confidence intervals, computed with bootstrap, are the following (we also provide the results from a one-sample t-test): (i) duration: $[0.175, 0.245]$ ($t(22) = 11.3$, $p =$ 1.3e−10), (ii) relative duration: $[0.086, 0.105]$ ($t(22) = 19.3$, $p =$ 2.7e−15), (iii) velocity: $[0.059, 0.140]$ ($t(22) = 4.6$, $p =$ 1.5e−4), and (iv) relative velocity: $[0.147, 0.200]$ ($t(22) = 12.6$, $p =$ 1.5e−11).

3.2 Live performed corpus

Similarly, Fig. 3 summarizes the results for the performed corpus. There is a significant positive correlation between information and (i) duration ($r = 0.04$, $p <$ 1e−100, $z = 28$), (ii) relative duration ($r = 0.03$, $p <$ 1e−100, $z = 21$), (iii) velocity ($r = 0.22$, $p <$ 1e−100, $z = 163$), and (iv) relative velocity ($r = 0.18$, $p <$ 1e−100, $z = 131$). Here, we can see that the relationship between information and duration is clearly of a U-shape (which is actually also present in the written corpus if one looks at some of the individual results), which explains the lower Spearman correlation coefficient compared to the previous case. One can also notice that the correlation between information and (absolute) velocity is stronger than in the sequenced corpus, which makes sense given that velocity is much more meaningful in the case of the performed corpus (as discussed in Section 2.2).

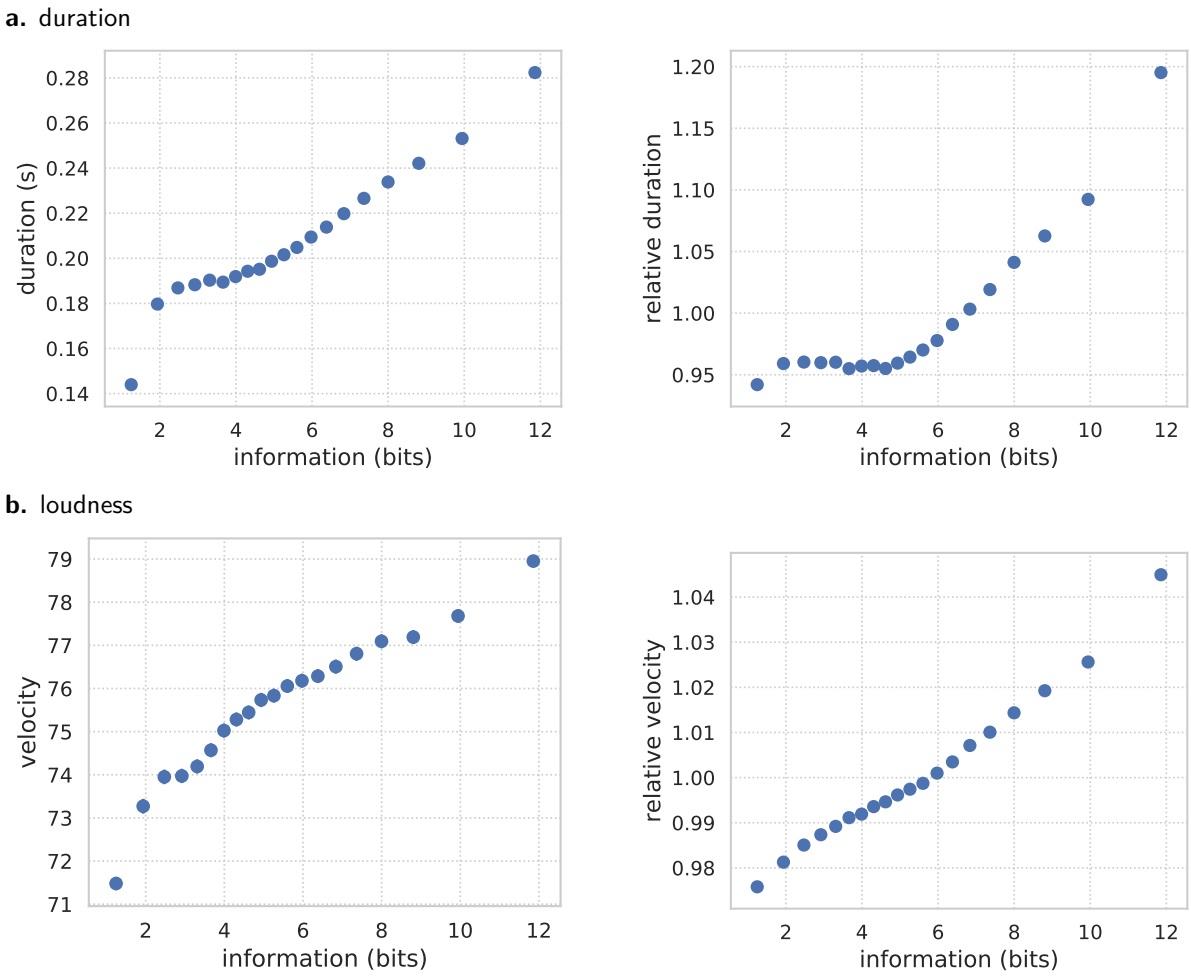

Figure 2: **Relationships between the information content of a harmonic segment and its duration (top) or its loudness (bottom), in the written corpus, pooling all composers together**. As a function of information content: (a, left) Absolute duration, in seconds. (a, right) Relative duration, defined as the ratio between the duration of current segment and the average duration of segments within a centered window of 10 s. (b, left) Absolute MIDI velocity, proportional to the loudness. (b, right) Relative velocity, defined as the ratio between the velocity of current segment and the average velocity of the 5 center-surrounding segments. Each bin represents 5% of the data. Error bars indicate 95% confidence intervals, estimated by bootstrap (a given bar might not be visible if below the size of its corresponding marker).

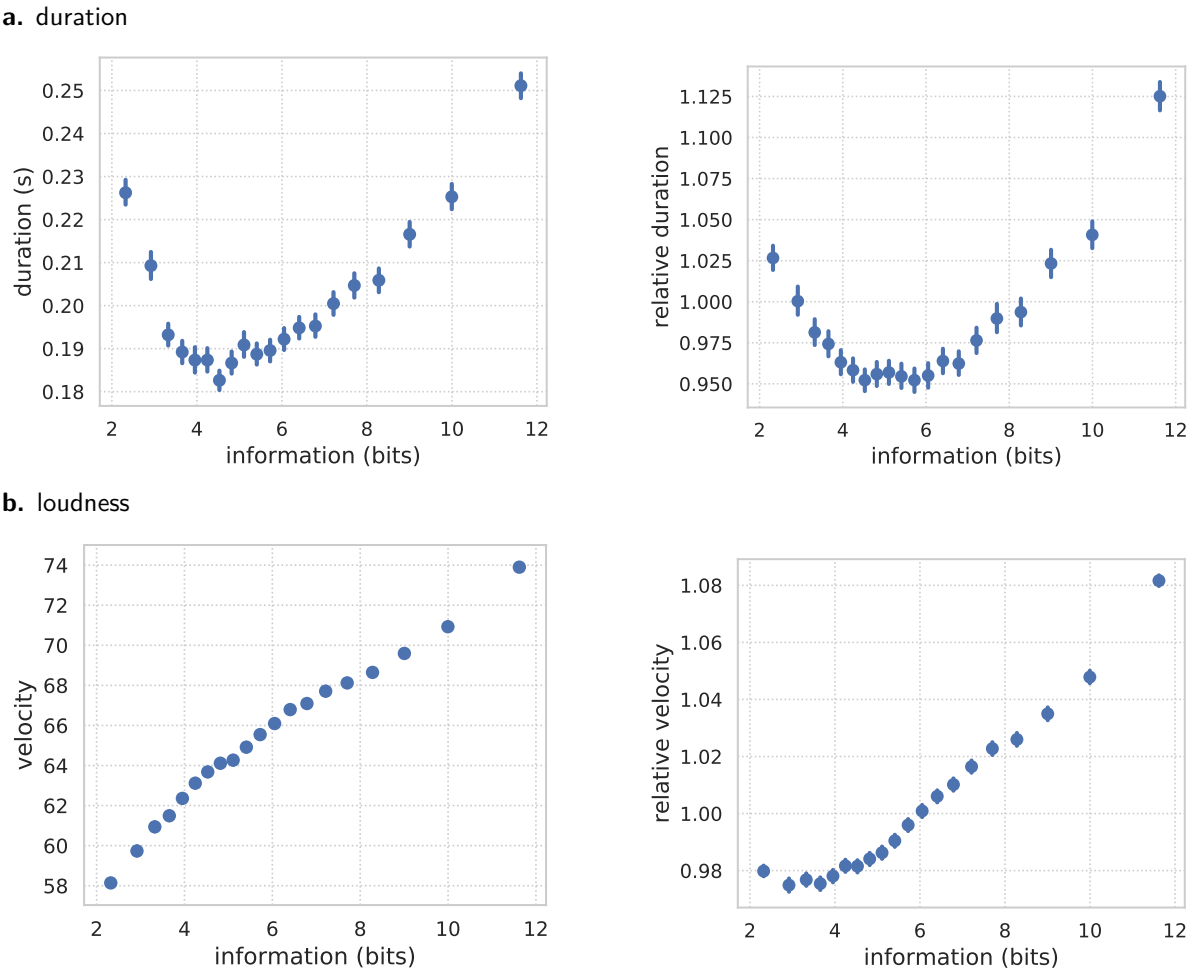

Figure 3: **Relationships between the information content of a harmonic segment and its duration (top) or its loudness (bottom), in the performed corpus**. Same legend as in Fig. 2.

3.3 Explaining the U-shape relationship between duration and information

In the case of duration, one can immediately notice that the relationship is actually more like a U-shape than a monotonic increase. Events with low information content tend to be associated with longer duration. This is clearly marked in the case of the performed corpus, and explains why the correlation is particularly weak in that case. This U-shape can be understood by the following phenomenon. Common practice period is usually characterized by stereotypical phrase ending, often following the same harmonic sequence, typically (variations of) the ii-V-I, subdominant, dominant, tonic progression. This should typically translate into a decrease in information at phrase and sectional boundaries. Final resolution was indeed experimentally shown to be characterized by a decline in information content (Manzara et al., 1992; Witten et al., 1994). At the same time, the ending of a musical piece is more than often performed with a final ritardando, *ie* a decrease in tempo (Sundberg and Verrillo, 1980; Desain and Honing, 1996; Friberg and Sundberg, 1999) (more generally, this phenomenon is also observed at the end of each phrase and each section, Windsor and Clarke, 1997). Here, simply looking at the ending of each MIDI file (hence without the need of referring to the notion of phrases or structures) and using our estimation of information content, we can look at the mean profile of surprisal: as expected, as music gets closer to its end, information typically decreases (Fig. 4a), while duration of each segment increases (Fig. 4b). If we restrict our analysis to the first half of the ending, where information and duration are still rather flat as a function of position, the relationship between duration and information is indeed

monotonic (Fig. 4c, left), whereas the U-shape is clearly marked if we consider the ending of the pieces, where information decreases and duration increases (Fig. 4c, right). A more detailed analysis confirms this visualization. For each of these two halves, and for each composer, we fitted a linear model aiming at explaining duration as a function of information and position from the end. Fig. 4d presents the distribution of the resulting regression coefficients. In both cases, information content is a significant predictor of duration (one-sample t-test: $t(22) = 17.2$, $p = 2.9\text{e}{-14}$ for the first half, $t(22) = 16.4$, $p = 8.4\text{e}{-14}$ for the second half), whereas the influence of position is much greater in the second half (one-sample t-test, considering the difference, $t(22) = 9.4$, $p = 3.6\text{e}{-9}$), where the U-shape is clear. Finally, for each MIDI file in our sequenced corpus, we computed a U-shapeness score by taking the quadratic coefficient from a second order polynomial fit. We also estimated the decrease in information by taking the ratio between the average of the last two values and the average value of the information from a more plateau region preceding it: values below 1 indicates a decrease in information, and the lower the value the sharper the decrease. Supplementary Fig. S6 shows that for values of this ending ratio below 1, there is a negative relationship with the U-shapeness score: the stronger the decline in information, the more U-shaped the relation between duration and information is.

All in all, the ending of a piece is therefore characterized by both a decrease in information and an increase in the average duration between two onsets, hence the U-shapeness of the relationship between duration and information. The typical final ritard as well as the decreases in tempo at phrase and sectional boundaries are a hallmark of expressive performances, which explains why the U-shape relationship is more marked in the case of performed music.

3.4 Analysis of a possible confound: musical meter

Harmonic rhythm is often aligned with meter, meaning that harmonic changes usually happen between bars, while chords tend to stay the same within a given bar. Meanwhile, it is also well known that events on strong beats are typically longer in duration (Palmer and Krumhansl, 1990; Longuet-Higgins and Lee, 1982). All in all, the observed association between duration and information could simply be due to these already well known effects of the metrical structure: a chord at the beginning of a measure tends to be both unexpected (informative) and long in duration (the same phenomenon actually apply for relationships between information and loudness, intensity being also used as a cue for meter, see e.g. Palmer and Krumhansl, 1990). First, note that even if the information/duration relationship were essentially driven by this metrical confound, it would still be compatible with the efficient communication framework. We have considered that informative items would be emphasized by means of phenomenal accents (duration and loudness here), but this could indeed also be realized thanks to metrical accents. It has been proposed that metrical prominence induces greater attention (Large and Jones, 1999; Large and Palmer, 2002), which means that the processing of unexpected chords would be eased by their prominent position in the metrical structure.

By taking metrical strength into account in our analysis, we show that the information/duration relationship actually holds beyond this possible metrical confound. We restrict ourselves to the written corpus, for which we have access to valid bar positions, and consider all the pieces with a 4/4 time signature. We model metrical hierarchy (Lerdahl and Jackendoff, 1985) by considering four levels of metrical prominence as a function of the position within a measure: events on the first beat in a bar receive the highest prominence, followed by those lying on beat 3, then those on beat 2 and 4, while events on any other temporal position (*ie* not falling on a beat) receive the lowest prominence (see Supplementary Fig. S7a). First, we find that the well-known above-mentioned associations hold, motivating our initial

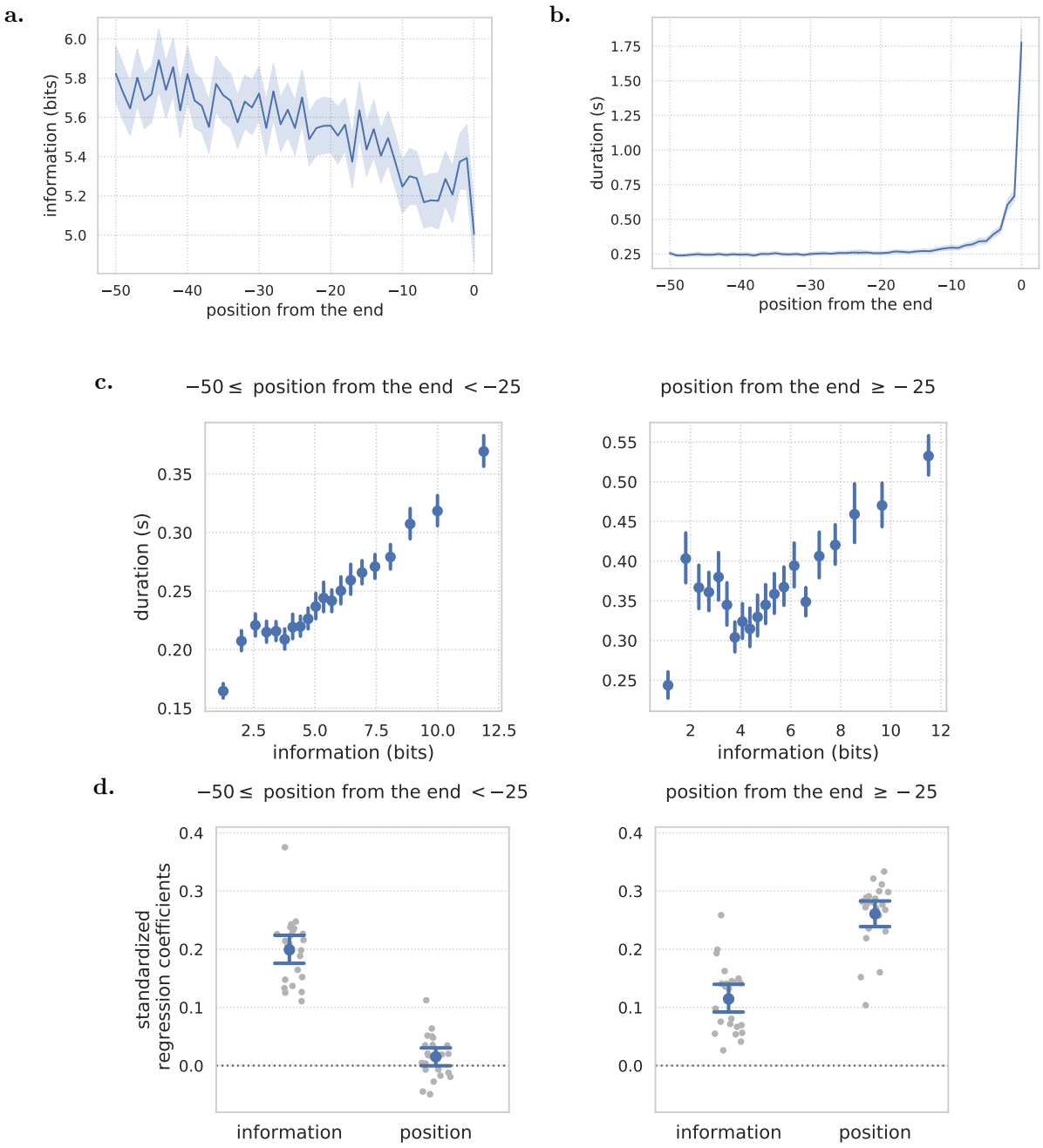

Figure 4: **Analysis of the ending of each piece, or why low information content can be associated with greater duration**. Looking at the ending (the last 50 segments) of each MIDI file of the sequenced corpus: (a) Mean information content (in bits) of a segment as a function of its position from the end. (b) Same with mean duration (in seconds). (c) Duration as a function of information, for the first half of the ending of each piece (left) and the second half (right). (d) Distribution of the standardized regression coefficients of a linear model '`duration ~ information + position`', computed for each composer in the sequenced corpus, for the first half of the ending of each piece (left) and the second half (right). The gray dots correspond to individual composers, while the blue dot represents the mean. Error bars indicate the 95% confidence intervals, estimated by bootstrap.

concern: the greater the metrical prominence, the longer the duration (Supplementary Fig. S7b), and the higher the information (Supplementary Fig. S7c). However, combining information and metrical prominence into a regression model that aims at explaining duration, we find that both quantities actually play a role (see statistical analysis in Supplementary Fig. S7d).

3.5 Diachronic changes

On a more minor note, the present study allows one to look at diachronic changes of the information content of music (here through the lens of harmonic flow). Computing different information-related quantities for each composer, and comparing them with the birth dates, we find that high information (defined as the 95th percentile), information rate (information per unit of time), and information breadth (defined as the difference between the 95th and the 5th percentile) increase over time (see Supplementary Fig. S8). As for the U-shapeness of the relationship between duration and information, we might expect to find such shape in the baroque, classical and romantic styles, but not so much in post-common practice music, for the tonal clichés got less and less prevalent. Looking at the detailed figures, and, for each composer, estimating a U-shapeness score as the quadratic coefficient of a second order polynomial fit, we can see that this is indeed the case to a certain extent, although we still find a clear U-shape relationship for some nineteenth-century composers (see Supplementary Figs. S1, S2 and S8).

4 Discussion and Conclusion

We found that the more surprising a harmonic event, the longer and the louder it will tend to be written and performed. The results of the present study suggest that efficient communication principles are at work in music, complementing the evidence already provided by the ample literature on language and speech.

Interestingly, in the case of duration, the relationship is more of a U-shape. This relates with the two kinds of positions that have been taken with respect to musical meaning. Following Meyer (1956), one view, the "absolutist", states that "musical meaning lies exclusively within the context of the work itself", whereas the other one, the "referentialist", asserts that "music also communicates meanings which in some way refer to the extramusical world of concepts, actions, emotional states, and character". Meyer also talks respectively about "embodied" and "designative" meanings. This author mainly focuses on the former aspect, while the latter view lies at the heart of the work of Schlenker (2016). As we have shown, musical pieces (and musical phrases) typically end with a decrease in both information and tempo. This slow down has been repeatedly interpreted as resulting from the association of music with a (possibly virtual) source that loses energy, thus decelerating (Sundberg and Verrillo, 1980; Desain and Honing, 1996; Schlenker, 2016). This phenomenon operates in addition to the positive relationship between information and duration that directly follows the efficient communication principle, which is only concerned by close internal connections that arise from the specific organization of the musical material itself.

A working hypothesis of our work assumes that music acts as some form of communication, at least in part. This point has been extensively discussed elsewhere (see e.g. Juslin and Laukka, 2003; Miell et al., 2005; Cross, 2014) and is not the topic of this study. Communication and information are loaded concepts, and should only be considered here from a Shannon information theoretic point of view. We do not say anything about the content of the messages that are communicated (their meaning), nor about why there is communication. For the sake of clarity, let us summarize this framework. We consider two entities, a source and a destination (here e.g. a composer and a listener, or a musician and an audience).

The source sends a message which is encoded into a certain signal. The destination has to decode it to reconstruct the original message with low probability of error, despite channel noise, using mutual knowledge shared by both parties. This common knowledge is crucial for the current discussion, and is in part constituted by the contextual probabilities that define a specific style. The present results show that communication is efficient in the sense that it adapts the resources involved in the process as a function of the information content of a segment, thus minimizing the effort in achieving successful transmission. Note that this 'minimization of effort' is somehow vague, and can be distributed along the whole chain of processing: for instance, it might reduce the physical effort from the sender (the sound-maker), or facilitate comprehension from the listener, notably by making perception more robust to noise, or by easing the memory process. Teasing out where and why exactly this efficiency occurs is a question that cannot be answered easily, and probably not with the current data.

Keeping a rather constant rate of complexity in information flow could have another additional benefit. It is well known that music lies at the edge between order and chaos, in the sense that it is neither too monotonous nor too disordered, and this property was recently shown to be a musical universal shared across different cultures (Mehr et al., 2019). By allocating more resources to more complex elements while reducing effort on the obvious, the efficiency principle could ensure that any event lies in a sweet spot of complexity, neither too simple nor too complex, following the Goldilocks Effect formulation that was found by Kidd et al. (2012): infants prefer to look at visual sequences that have a complexity that is neither too high nor too low, yielding a U-shape relationship between looking time and complexity. Here, the efficiency principle might help keep the musical content not only understandable, but also interesting.

The present results do not say anything a priori about aesthetics, but simply show how efficient communication shapes both musical composition and performance. Still, the notion of information depends on a particular model, which is built on previous musical exposure. Proper understanding of a style, appreciation of the subtleties of a particular musical interpretation requires learning, *ie* enculturation, which might then influence liking. The present framework might thus help to investigate the role of efficient communication with respect to aesthetic judgments and musical manifestation (either composed or performed), following Meyer (1956), who claims that playing with the expectancies created by the musical work are at the heart of the musical emotions.

Many directions would be worth exploring in future work. First, we might investigate other means to give prominence to a musical segment. We have considered duration and intensity (although we have also briefly considered the role of metric prominence and attentional resources in section 3.4), which are important dimensions of musical expressiveness, but many other accents could also be considered: agogic accents, melodic accents by means of skips or turns, changes in articulation, to name but a few (see Parncutt, 2003, for a catalog). A composer or a player has access to the whole palette – how these elements are combined to provide the right amount of prominence depending on the information content remains an open empirical question. Second, we have only considered predictability from the point of view of using the past to predict the future. In other words, following Western music notation, flowing in a left to right manner, we used the left context of a given segment. But the case that reverses time would also be an interesting perspective: the context on the right side of a given element might explain *a posteriori* the presence of this element, as illustrated by the following quote, attributed to famous jazz pianist and composer Bill Evans: "There are no wrong notes, only wrong resolutions."[5]. Finally, other aspects of music, notably melodic and rhythmic, and other musical genres beyond Western music will be

[5]See e.g. `https://commonreader.wustl.edu/c/the-art-of-the-mistake/`

worth investigating in the future.

Acknowledgments

Many thanks to Jean-Pierre Nadal, Christophe Pallier, Emmanuel Chemla and Philippe Schlenker for helpful comments and valuable discussions on an earlier version of this work. I would also like to thank Florian Jaeger, Richard P. Cooper and three anonymous reviewers for their detailed and insightful comments and suggestions.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, 47(1):31–56.

Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119(5):3048–3058.

Bartlette, C. A. (2007). *A Study of Harmonic Distance and Its Role in Musical Performance*. PhD thesis, University of Rochester.

Bigand, E., Madurell, F., Tillmann, B., and Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1):184.

Bonnasse-Gahot, L. (2019). Data and code for "Efficient communication in written and performed music". `https://osf.io/gxw4b`, doi:10.17605/OSF.IO/GXW4B.

Boulanger-lewandowski, N., Bengio, Y., and Vincent, P. (2012). Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. In *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland.

Brooks, F. P., Hopkins, A., Neumann, P. G., and Wright, W. V. (1957). An experiment in musical composition. *IRE Transactions on Electronic Computers*, (3):175–182.

Carlson, R., Friberg, A., Frydén, L., Granström, B., and Sundberg, J. (1989). Speech and music performance: Parallels and contrasts. *Contemporary Music Review*, 4(1):391–404.

Cherry, E. C., Halle, M., and Jakobson, R. (1953). Toward the logical description of languages in their phonemic aspect. *Language*, pages 34–46.

Chollet, F. et al. (2015). Keras. `https://keras.io`.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, 7(2):137–163.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, 6(2):243–278.

Cohen Priva, U. and Jurafsky, D. (2008). Phone information content influences phone duration. In *Conference on prosody and language processing*.

Cover, T. M. and Thomas, J. A. (2006). Elements of information theory 2nd edition. *Willey-Interscience: NJ*.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music*, 42(6):809–819.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray, London.

Desain, P. and Honing, H. (1996). Physical motion as a metaphor for timing in music: the final ritard. In *Proceedings of the 1996 International Computer Music Conference*, pages 458–460. ICMA.

Deutsch, D. (1980). Music perception. *The Musical Quarterly*, 66(2):165–179.

Dubnov, S., Assayag, G., Lartillot, O., and Bejerano, G. (2003). Using machine-learning methods for musical style modeling. *Computer*, (10):73–80.

Eck, D. and Schmidhuber, J. (2002). Learning the long-term structure of the blues. In *International Conference on Artificial Neural Networks*, pages 284–289. Springer.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2):179–211.

Fedorenko, E., Behr, M. K., and Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108(39):16428–16433.

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., and Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, 108(12):3289–3300.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Memory & cognition*, 37(1):1–9.

Fenk, A. and Fenk, G. (1980). Konstanz im kurzzeitgedächtnis-konstanz im sprachlichen informationsfluß. *Zeitschrift für experimentelle und angewandte Psychologie*, 27:400–414.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. *Typological Studies in Language*, 45:431–448.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1):173–215.

Frank, A. F. and Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 30.

Friberg, A. and Sundberg, J. (1999). Does music performance allude to locomotion? a model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America*, 105(3):1469–1484.

Futrell, R. and Levy, R. P. (2019). Do rnns learn human-like abstract word order preferences? *Proceedings of the Society for Computation in Linguistics (SCiL)*, pages 50–59.

Gal, Y. and Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. In *Advances in neural information processing systems*, pages 1019–1027.

Genzel, D. and Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 199–206.

Gers, F. A. and Schmidhuber, E. (2001). Lstm recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks*, 12(6):1333–1340.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.

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

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1):23–62.

Jeong, D. and Nam, J. (2017). Note intensity estimation of piano recordings by score-informed nmf. In *Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Audio Engineering Society.

Juslin, P. N. and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, 129(5):770.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, 7(5):e36399.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, 2(15):207–212.

Koelsch, S., Rohrmeier, M., Torrecuso, R., and Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*, 110(38):15443–15448.

Krumhansl, C. L. and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review*, 89(4):334.

Large, E. and Jones, M. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1):119–159.

Large, E. W. and Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive science*, 26(1):1–37.

Lerdahl, F. (2004). *Tonal pitch space*. Oxford University Press.

Lerdahl, F. and Jackendoff, R. S. (1985). *A generative theory of tonal music*. MIT press.

Levy, R. P. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems*, pages 849–856.

Linzen, T., Dupoux, E., and Goldberg, Y. (2016). Assessing the ability of lstms to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.

Longuet-Higgins, H. C. and Lee, C. S. (1982). The perception of musical rhythms. *Perception*, 11(2):115–128.

Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., and Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318.

Mandelbrot, B. (1954). Structure formelle des textes et communication. *Word*, 10(1):1–27.

Manin, D. Y. (2006). Experiments on predictability of word in context and information rate in natural language. *Journal of Information Processes*, (3):229–236.

Manzara, L. C., Witten, I. H., and James, M. (1992). On the entropy of music: An experiment with Bach chorale melodies. *Leonardo Music Journal*, 2(1):81–88.

Markov, A. A. (1913). Essai d'une recherche statistique sur le texte du roman 'eugène oneguine', illustrant la liaison des épreuves en chaîne. *Bulletin de l'Académie Imperiale des Sciences de St. Pétersbourg*, 7(3):156–162.

McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., et al. (2019). Universality and diversity in human song. *Science*, 366(6468).

Melis, G., Dyer, C., and Blunsom, P. (2018). On the state of the art of evaluation in neural language models. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*.

Merity, S., Keskar, N. S., and Socher, R. (2017). Regularizing and optimizing lstm language models. *arXiv preprint arXiv:1708.02182*.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: Chicago University Press.

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, 15(4):412–424.

Miell, D., Macdonald, R., and Hargreaves, D., editors (2005). *Musical communication*. Oxford University Press.

Palmer, C. and Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4):728.

Parncutt, R. (2003). Accents and expression in piano performance. *Perspektiven und Methoden einer Systemischen Musikwissenschaft*, pages 163–185.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, 6(7):674.

Pearce, M. T. and Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Topics in cognitive science*, 4(4):625–652.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Peretz, I., Vuvan, D., Lagrois, M.-É., and Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1664):20140090.

Piantadosi, S. T., Tily, H., and Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.

Piantadosi, S. T., Tily, H. J., and Gibson, E. (2009). The communicative lexicon hypothesis. In *The 31st annual meeting of the Cognitive Science Society (CogSci09)*, pages 2582–2587.

Pinkerton, R. C. (1956). Information theory and melody. *Scientific American*, 194(2):77–87.

Raffel, C. and Ellis, D. P. (2014). Intuitive analysis, creation and manipulation of midi data with pretty_midi. In *15th International Conference on Music Information Retrieval Late Breaking and Demo Papers*, pages 84–93.

Rohrmeier, M. A. and Koelsch, S. (2012). Predictive information processing in music cognition. a critical review. *International Journal of Psychophysiology*, 83(2):164–175.

Schlenker, P. (2016). Prolegomena to music semantics. *Review of Philosophy and Psychology*, pages 1–77.

Schoenberg, A. (1978). *Theory of harmony*. University of California Press, Berkeley, Los Angeles. translated by Roy E. Carter.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell system technical journal*, 30(1):50–64.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Sundberg, J. and Verrillo, V. (1980). On the anatomy of the retard: A study of timing in music. *The Journal of the Acoustical Society of America*, 68(3):772–779.

Temperley, D. (2014). Information flow and repetition in music. *Journal of Music Theory*, 58(2):155–178.

Temperley, D. and Gildea, D. (2015). Information density and syntactic repetition. *Cognitive science*, 39(8):1802–1823.

Tillmann, B. (2012). Music and language perception: expectations, structural integration, and cognitive sequencing. *Topics in cognitive science*, 4(4):568–584.

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22.

Walder, C. (2016). Modelling symbolic music: Beyond the piano roll. In *Asian Conference on Machine Learning*, pages 174–189.

White, C. W. and Quinn, I. (2016). The yale-classical archives corpus. *Empirical Musicology Review*, 11(1).

Windsor, W. L. and Clarke, E. F. (1997). Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool. *Music Perception: An Interdisciplinary Journal*, 15(2):127–152.

Witten, I. H., Manzara, L. C., and Conklin, D. (1994). Comparing human and computational models of music prediction. *Computer Music Journal*, 18(1):70–80.

Youngblood, J. E. (1958). Style as information. *Journal of Music Theory*, 2(1):24–35.

Zipf, G. K. (1936). *The psycho-biology of language: An introduction to dynamic philology*. Routledge.