journal of interdisciplinary music studies
2014-2016, volume 8, issue 1&2, art. #16081204, pp. 91-120
open peer commentary article

Music perception in historical audiences: Towards predictive models of music perception in historical audiences

Marcus T. Pearce[1] and Tuomas Eerola[2]

[1]Queen Mary University of London, London E1 4NS
[2]Durham University, Durham DH1 3RL

Background in Historical Musicology. In addition to making inferences about historical performance practice, it is interesting to ask questions about the experience of historical listeners. In particular, how might their perception vary from that of present-day listeners (and listeners at other time points, more generally) as a function of the music to which they were exposed throughout their lives.
Background in Music Cognition. To illustrate the approach, we focus on the cognitive process of expectation, which has long been of interest to musicians and music psychologists, partly because it is thought to be one of the processes supporting the induction of emotion by music. Recent work has established models of expectation based on probabilistic learning of statistical regularities in the music to which an individual is exposed. This raises the possibility of developing simulations of historical listeners by training models on the music to which they might have been exposed.
Aims. First, we aim to develop a framework for creating and testing simulated perceptual models of historical listeners. Second, we aim to provide simple but concrete illustrations of how the simulations can be applied in a preliminary approach. These are intended as illustrative feasibility studies to provide a springboard for further discussion and development rather than fully fledged experiments in their own right. Third, we aim to appeal to the expertise of historical musicologists in identifying useful research questions and appropriate constraints for the simulations, so these can be used to complement existing evidence on the perception of music by historical listeners.
Main contribution. Our primary contribution is to develop and illustrate a framework which we believe can shed light on the perception of music by historical listeners and, in particular, how listeners of different periods might have generated different predictions to music as a function of differences in their musical experiences. The framework we develop involves several steps. First, identifying a research question; second, selecting a corpus (or corpora) to represent the musical experience of the listener(s) we want to simulate; third, identify the central musical features of interest and use them to develop a representation scheme for the selected compositions; finally, the model parameters are selected and the models are trained on the selected corpora to simulate particular listeners. We identify and discuss the decisions that must be made at each step. Finally, we illustrate the framework by training models on a range of corpora from different stylistic traditions from different locations and points in history, including analyses at the level of entire collections, individual compositions, and individual events.
Implications. The results of our illustrative analyses suggest that the trained models behave as we hypothesised, demonstrating sensitivity to stylistic similarities which could illuminate how listeners from different eras might have experienced musical structures. However, the approach is in need of expertise in historical musicology to establish clear and relevant research questions and to select appropriate parameters for the simulations. With such additional input, we believe simulated listeners will provide important insights, alongside other evidence, into the question of how our forebears experienced the music of their time.
Keywords: music cognition, machine learning, historical musicology.

• *Correspondence:* Dr Marcus T. Pearce, School of Electronic Engineering & Computer Science, Queen Mary University of London, E1 4NS, UK, tel: +44 (0) 20 7882 6207, e-mail: marcus.pearce@qmul.ac.uk
• *Received:* 3 December 2013; *Revised:* 4 February 2015; *Accepted:* 7 September 2015
• *Available online:* 15 March 2017
• doi: 10.4407/jims.2016.12.004

1. Introduction

Recent decades have seen an increasing interest in historically informed performance, in which attempts are made to follow period performance practice wherever possible. However, the musical perception of modern audiences is affected by the music they have listened to throughout their lives and this music, in turn, will, in most cases, reflect several hundred years of musical change (e.g. Butt, 2002; Leech-Wilkinson, 2002). Of course, we cannot recreate a historically authentic audience (at least not ethically). But perhaps we can simulate one.

Computational methods provide for the intriguing possibility of creating models that simulate the perception of listeners of different periods. In particular, we propose that machine learning methods for modelling the perception and cognition of music can be used to simulate the perception of an historical audience by training the model on the music that a listener might have heard during their lifetime.

The approach may be capable of speaking to many interesting questions about early music itself as well as how it was perceived by contemporary listeners. One example is pitch spelling (e.g., Knopke & Jürgensen, 2012; Meredith, 2006) in which a model of how music was perceived could help to identify ambiguous pitch spellings in historical sources. It could also be useful in contributing evidence in cases of ambiguous authorship (e.g., van Kranenburg, 2008) and questions of stylistic influence between composers and performers (see, e.g., Cook, 2007, for an example in performance). On the level of musical style analysis, a model of a historical listener could make contributions in testing musicological theories about the characteristic features of musical styles (e.g., Volk & de Haas, 2013) and how such differences may be empirically explored by exposing contemporary listeners to them (Jürgensen, Pearson, & Knopke, 2014-2016). The approach could also contribute to questions of relevance in music cognition such as the development of tonal perception (e.g., Huron & Veltman, 2006).

This is a first attempt to define a framework for exploring the approach and we also give some illustrative examples as a very preliminary step towards illustrating its utility. The examples are just that – they are not intended to be fully fledged studies but to provide a proof of concept to inspire further discussion and development of the approach. In these examples, we focus on the cognitive process of expectation which has been widely studied in music cognition research (e.g., Huron, 2006; Meyer, 1956; Narmour, 1990), because it is an area in which we have experience (Eerola, 2004; Pearce, 2005) and which lends itself to a machine-learning approach. Furthermore, there is evidence that listeners' musical expectations are influenced by musical experience (Eerola, 2004; Narmour, 1990). However, the approach itself is general and should apply beyond expectation to other psychological processes in music perception (e.g., perceptual discrimination, similarity, classification, memory, emotion, attention). In so doing, it would be appropriate to model both aspects of music perception that are sensitive to musical experience and those that are not. We focus on the former category in our examples below because they describe phenomena that might change between historical musical periods.

The approach is also broadly compatible with corpus-based approaches in empirical musicology (e.g., Cook, 2007; Knopke & Jürgensen, 2012; Meredith, 2006; van Kranenburg, 2008; Volk & de Haas, 2013), with an additional focus on models that are psychologically motivated in the sense that they account well for listeners' perception of music. Finally, and importantly, to fully realise its potential, the approach needs interdisciplinary collaboration with historical musicologists to identify questions of interest and define appropriate constraints for the models.

2. A cognitive model of music perception

2.1. Introduction

In this section, we summarise a computational approach to modelling expectation in music perception (Pearce, 2005). Although this is only one of several models of expectation in the literature (e.g., Collins et al., 2014; Margulis, 2005; Milne et al., 2011; Narmour, 1990, Schellenberg, 1997; Temperley, 2007; Toiviainen & Krumhansl, 2003) and related concepts such as tension (e.g., Farbood, 2012; Lerdahl & Krumhansl, 2007; Rohrmeier, 2011), we choose it to exemplify the approach because it has useful features such as incorporating an experience-driven aspect, combining local (intra-opus) and longer-term (extra-opus) effects on musical expectation, and the ability to combine information from multiple musical features, including tonal and non-tonal effects (all of which are described further below). The purpose of the models described here is to understand the cognitive processes involved in generating expectations about forthcoming events while listening to music. Expectations are of general interest in psychology, but have particular relevance to music as they are thought to be involved in the induction of emotion in the listener (Huron, 2006; Juslin & Västfjäll, 2008; Meyer, 1956). In simulating this aspect of music perception, the task for the model is to predict some feature of the next event in the music. Here we use melodic pitch expectations as an example (i.e., predicting the pitch of the next note in a melody) but the approach generalises naturally to predicting the interval, timing or duration of the next note, combinations of features of the next note, the next chord in a sequence and so on.

2.2. Markov models

The central feature of the modelling approach is to learn sequential, statistical dependencies between notes in an unsupervised manner through exposure to melodies. This is achieved using Markov models or *n*-gram models (Manning & Schütze, 1999, ch. 9) viewing a melody as a sequence of non-overlapping events, each represented by a property such as pitch. An *n*-gram model computes the conditional probability of an event given the *n*-1 preceding events in the sequence. The quantity *n*-1 is called the *order* of the model. An *n*-gram model estimates the conditional probability of an event *e* given the context *c* of the preceding *n*-1 events based on the frequency with which that event occurred in the same context in the prior experience of the model.

$$p(e|c) = \frac{count(ce)}{count(c)} \quad (1)$$

where *ce* denotes the concatenation of *c* and *e*. The conditional probability $p(e|c)$ varies between 0 and 1. Given a trained *n*-gram model, the degree to which an event appearing in a given context is unexpected can be defined as the *information content* (IC), $h(e|c)$, of the event given the context and the model:

$$h(e|c) = \log_2 \frac{1}{p(e|c)}. \quad (2)$$

IC ranges from 0 to infinity: low values correspond to high probability events and vice versa. Given an alphabet $\mathcal{E}$ of events (e.g., a set of pitches in our case) which have appeared in the prior experience of the model, the uncertainty of the model's expectations in a given melodic context can be defined as the *entropy,* or average information content, of the events in $\mathcal{E}$:

$$H(c) = \sum_{e \in \mathcal{E}} p(e|c) h(e|c). \quad (3)$$

Entropy ranges from 0 to $\log_2(|\mathcal{E}|)$: low values indicate low uncertainty and vice versa. In modelling musical perception, entropy represents the predictive uncertainty about which musical event will come next in a given context (before that note actually arrives) while information content reflects the unexpectedness of the event that actually does follow.

2.3. Extensions to Markov Models

The present framework extends basic Markov modelling in three ways. First, we consider models with different fixed orders. Variable-order models, where the order used varies throughout prediction (Cleary & Teahan, 1997; Pearce, 2005) are also possible but not considered here.

Second, the system may be configured with two components (Conklin & Witten, 1995; Pearce et al., 2006): first, the long-term model (LTM), which is exposed to an entire corpus (representing schematic effects of long-term exposure to music); second, a short-term model (STM), which is exposed only to the current melody (representing short-term processing of local structure in the current listening episode). The models may be used in combination or in isolation. There are five configurations: the STM alone (STM); the LTM alone (LTM); the LTM+, which is a version of the LTM that learns dynamically while predicting, in comparison to the LTM which is fixed and static after training; the STM and LTM together (BOTH); and the STM and LTM+ together (BOTH+). When used in combination (BOTH, BOTH+), each model generates a probability distribution predicting the pitch of each note as the melody proceeds, which are then combined (Conklin & Witten, 1995, Pearce et al., 2005). In the examples provided below, we focus on the long-term model (LTM). The reasons for this are explained in 3.1.1.

Third, the framework allows for modelling and combining different features present in and derived from the events making up the musical surface. We have to treat music, and the perception of it, as a multidimensional phenomenon since musical elements differ in pitch, timing, loudness, timbre, spatial location and other attributes and these dimensions are known to influence perception (Levitin & Tirovalas, 2009). Pitch perception alone shows evidence of multidimensional cognitive representations (Shepard, 1982). Therefore, each note is represented as a discrete event consisting of a conjunction of basic features such as pitch, onset time, duration, loudness etc., which may assume one of a finite set of values (the *alphabet* or *domain* of that feature). A *multiple viewpoint framework* (Conklin & Witten, 1995; Pearce et al., 2005) may be used to predict a basic feature (e.g., pitch, onset time) using multiple models trained on different abstract derivations of the basic feature (e.g., scale degree, inter-onset interval). Furthermore, such representations may be selected automatically to improve the predictive accuracy of the model by minimising information content (Chater, 1999; Pearce, 2005). In this preliminary work, we do not use these advanced features of the framework, focusing instead on comparing two basic features: scale degree (MIDI note number mod 12, relative to tonal centre) and pitch interval (from previous note, in semitones, with sign representing direction).

2.4. Summary

Using the system involves choosing a configuration (i.e., STM, LTM, LTM+, BOTH, BOTH+), choosing the basic *target viewpoints of interest* and then choosing the set of *source viewpoints* used in prediction (either manually or using viewpoint selection).

In this work we compare long-term models (LTM) with varying fixed orders using target viewpoints of scale degree and pitch interval (in this work the source viewpoint is always the same as the target). Future work will develop more sophisticated models of historical listeners using more complex sets of modelling parameters.

We focus on Markov models since they have been shown to be capable of learning stylistic characteristics and accurately predicting listeners' expectations. However, there is also work on other kinds of machine learning methods for music prediction, including neural networks (e.g., Cherla et al., 2013). Although the application of these methods to musical structure is in its infancy, they could be substituted into the framework in future, should they be shown to model expectations more accurately than the Markov models used in our examples below.

2.5. Empirical support for the framework

The use of this approach in a cognitive model of auditory expectation is motivated by empirical evidence of implicit learning of statistical regularities in musical melody and other sequences of pitched events (Oram & Cuddy, 1995; Saffran et al., 1999). Consistent with a process of statistical learning, melodic pitch expectations vary between musical styles (Krumhansl et al., 2000) and cultures (Carlsen, 1981;

Castellano et al., 1984; Curtis & Bharucha, 2009; Eerola, 2004, Kessler et al., 1984; Krumhansl et al., 1999), throughout development (Schellenberg et al., 2002) and across degrees of musical training and familiarity (Krumhansl et al., 2000; Pearce et al., 2010). In particular, research exploring differences in expectation between listeners with experience of different musical cultures (Carlsen, 1981; Castellano et al., 1984; Curtis & Bharucha, 2009; Eerola, 2004, Kessler et al., 1984; Krumhansl et al., 1999) is complementary to our proposal to study differences in expectation between listeners with experience of music of different historical periods. Indeed, recent work has attempted to simulate these effects of cultural musical exposure, using an approach related to the one we propose here (Curtis & Bharucha, 2009).

In particular, there is evidence that pitch expectations are informed both by long-term exposure to music and by the encoding of regularities in the immediate context. Krumhansl, for example, showed that tonal expectations derived from probe-tone experiments (Krumhansl & Kessler, 1982) are closely related to zeroth-order distributions of chromatic scale degrees in large collections of music. Krumhansl argued that tonal hierarchies are acquired by statistical learning through long-term exposure to music. There is also evidence that local musical structure influences expectations. Oram & Cuddy (1995) conducted a series of experiments in which continuation tones were rated for musical fit in the context of artificially constructed sequences of pure tones in which the tone frequencies were carefully controlled. The continuation tone ratings of both trained and untrained listeners were significantly related to the frequency of occurrence of the continuation tone in the context sequence. Tillmann and colleagues have shown that target chords are processed more accurately and quickly when they are related both to the local and the global harmonic context (previous chord and prior context of six chords) respectively (Tillmann et al., 1998) and that these effects can be explained by a mechanism of implicit statistical learning of sequential harmonic patterns in music (Tillmann et al., 2000).

There is also evidence that pitch expectations are influenced by higher-order probabilistic prediction. Saffran et al., (1999) showed that infants and adults are capable of implicitly learning first-order probabilities in tone sequences and using them to identify segment boundaries. These influences also hold for musical stimuli. In a study using Finnish spiritual hymns, Krumhansl, Louhivuori, Toiviainen, Järvinen & Eerola (2001) presented evidence for the influence of second-order probabilities on listeners' pitch expectations. Extending these results, the model presented above has been tested by comparing its pitch expectations with those of human listeners (Pearce, 2005). In a series of reanalyses of existing behavioural data (Cuddy & Lunny, 1995; Manzara et al., 1992; Schellenberg, 1997), it was shown that this model predicts listeners' expectations better than existing models of melodic expectation based on innate principles (Narmour, 1990; Schellenberg, 1997). Using a novel visual cueing paradigm for eliciting auditory expectations without pausing playback, Pearce et al., (2010) confirmed that the model predicts listener's pitch expectations in melodies without explicit rhythmic structure. Recent work has extended these findings to entropy as a model of uncertainty in music perception (Hansen et al., 2013).

To summarise, many studies suggest that melodic expectancy is a strong candidate for tapping into acquired knowledge of music through exposure (for a review, see Tillmann, Poulin-Charronnat, Bigand, 2014). Rather than compare expectations between cultures, as done in previous research, we propose to examine expectations between historical periods. The obvious problem is that we no longer have access to listeners from those periods for empirical psychological research. However, attempts to simulate the sensitivities of listeners of a particular period have been conducted by giving contemporary listeners exposure to different historical styles. Jürgensen, Pearson, & Knopke (2014-2016), for example, explicitly investigated perception of historical change – where highly conventionalised treatment of dissonance gave way to unprepared dissonance in the early decade of 1600 – by exposing contemporary listeners to varying amounts of musical examples from the time periods involved. The results suggest that statistical learning through experience could account for some differences in familiarity ratings.

3. Methodology

3.1. Identifying Research Questions

The most important part of this proposal is to demonstrate the kind of research questions that can be addressed with cognitive modelling and how such an approach is connected to information derived from a diverse set of music corpora. Given the interdisciplinary nature of the endeavour, these must be coherent in terms of musicological value, psychological validity and computational feasibility.

How would a seventeenth-century listener have responded to a particular note in a Monteverdi madrigal? Are there events that would have been surprising to such a listener but which are not to a modern listener and vice versa? Would the sense of tension arising from a perception of uncertainty have been the same for a contemporary listener as for a modern day listener? These questions can be addressed to some extent by simulating listeners of different periods by exposing the model to music of those periods and comparing the responses of these simulated listeners to music of different periods (their own, future and past periods). Other research questions might address the amount of information present in different voices, the features enabling optimal prediction of music at different periods, issues of stylistic development and ambiguities involved in transcribing early sources.

3.1.1. The present research question

As a first step, we ask whether long-term models trained on different stylistic corpora (across cultures and time) are capable of simulating enculturated listeners to those corpora. Our preliminary approach is to compare models trained on different corpora, representing enculturated listeners in the respective stylistic traditions, to see if they differ in plausible ways. Specifically, we take a collection of seventeenth-century madrigals as our point in history and examine whether long-term models trained on increasingly distant collections of music (culturally and historically) make increasingly different predictions about the structure of the madrigals.

3.2. Selecting the Corpus

To simulate a given listener, the model needs to be exposed to the same music that the listener has heard throughout their lifetime. Ideally, we would also simulate issues affecting the encoding (e.g., attention, arousal, interest, order effects and so on) and retrieval of the music (e.g., memory limitations). Although it might be interesting to model a specific listener in this way, we propose as a starting point, to simulate a *typical* listener of a given period by training the model on a *representative* sample of the music available within the culture of that listener.

This raises both the thorny issue of what we mean by typical and representative and the vexed question of whether it is actually possible to construct such a corpus for a given historical period and culture. There is a balance to be struck between making the corpus so specific (in terms of modelling a given listener in a given place at a given time, to the extent that this is possible) that the results have limited scope and making it so general as to be meaningless. The appropriate approach will depend to some extent on the question being addressed. Another issue is that a representative sample of music for a typical listener of a given period in a given location may well have included pieces that have not survived or attracted continued interest and, therefore, are no longer readily available.

3.2.1. The present corpus

Here we illustrate the approach by comparing a seventeenth-century collection of madrigals[i] with other collections of monophonic vocal music, representing a continuum from related to unrelated musical genres. The aim is to explore how sensitive the various representations are in quantifying broad stylistic differences between the corpora at the level of entire collections, individual compositions and specific events.

3.2.2. Musical examples from the Coppini collection

We took examples from Aquilino Coppini's collection of madrigals titled *musica tolta da i madrigali* from 1607 (Jacobsen, 1998, 2003), which consist of 24 polyphonic vocal works by Banchieri, Gabrieli, Giovanelli, Marenzio, Monteverdi, Nanino and Vecchi. Coppini was an associate of Monteverdi and edited at least three books of madrigals (Rorke, 1984) under the support of the Cardinal Borromeo (Macy, 2011). The advantage of this *Coppini collection* is that it is available in electronic format, both from KernScores[ii], and from the International Music Score Library Project[iii].

3.2.3. Derived long-term models of musical styles (LTMs)

To demonstrate the possibilities of the computational analysis of historical styles, a few points of reference are needed. In this case, an illustrative range of references is used to convey the sensitivity and consistency of the measures used. We chose six samples of music that vary in terms of stylistic similarity to the music contained in the *Coppini collection*.

The first collection comes from the substantial repertoire of J.S. *Bach chorales*, collected by C. P. E. Bach after J.S. Bach's death (Dörffel, 1940). These 371 chorales are available in KernScores[iv].

Second, we were fortunate to have access to an *Early Renaissance collection* containing 384 polyphonic secular works composed between 1350-1450, many from the *Chantilly Manuscript* (Reaney, 1954). This corpus has been curated, encoded and supplied by Michael W. Beauvois.

A more distant collection of vocal music is taken from a small sample of lieder by Franz Schubert (35 Lieder) and songs by Stephen Foster (38 songs). This collection, referred to here as *Schubert & Foster songs* (vocal lines only), is available from KernScores[v].

Another stylistically distant collection of songs from KernScores is contained in the *Essen collection* (Schaffrath, 1995), from which we take all German folk songs (5365). Again, these are readily available in Kern format[vi].

For our next most distant frame of reference, we took a sample of popular music, *UK top 10 hits* between 1960 and 1975, which contains 484 songs. We refer to this collection as *Pop songs*.

Finally, the most remote point of comparison is taken from *Native American songs* of the Ojibway, Sioux, and Pawnee collected by Frances Densmore. This corpus of 366 songs, edited by Paul von Hippel, is available online[vii], and serves to illustrate the farthest departure from the original renaissance sample. By way of summary, basic information about these collections is presented in Table **Error! Reference source not found.**.

Table 1. Summary of the collections.

Collection	N	Voices	Notes	Source
Coppini collection	24	4-5	3112	KernScores
Bach chorales	371	4	84666	KernScores
Early Renaissance collection	384	4-5	51236	Private
Schubert & Foster songs	73	Monophonic	10635	KernScores
Essen collection	5365	Monophonic	301617	KernScores
Pop songs	484	Monophonic	131697	Private
Native American songs	366	Monophonic	22740	KernScores

3.3. Selecting the representation

Once a suitable corpus has been identified, we have to ask how best to represent the music for the model. Although we are interested in modelling human perception, particularly melodic expectations, we do not attempt a completely accurate reconstruction of the physiological and psychological processes involved in auditory representation of sound from the ear upwards. Rather, we identify a level of representational abstraction in auditory processing that is appropriate to the questions of interest. Here, we are concerned with high-level musical structure, so we assume that lower-level mechanisms deliver a note-like representation of music, and take this as our musical surface.

3.3.1. Transcription

There are many issues involved in making accurate transcriptions into modern staff notation of original sources, whose notation systems may require interpretation and which may be handwritten, incomplete, damaged or otherwise ambiguous in a variety of ways. One potential application of the approach presented here would be to use a trained model to disambiguate the process by, for example, making predictions about the most probable pitch spelling or note duration (Knopke & Jürgensen, 2012).

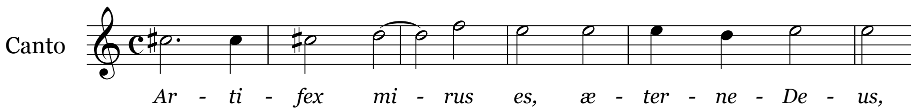
3.3.2. Basic representations

It makes sense to start with the most fundamental properties of notes: their pitch and timing (duration and onset time), though other features such as dynamics, timbre and articulation would also be of interest. Therefore the task set for the model would be to predict the onset time and absolute pitch (the former represented in beats, the latter represented as MIDI note numbers, see Table **Error! Reference source not found.**) of the next note in the music, given the previous notes (assuming for now that we are dealing with monody - we come to polyphony below). In our example below we focus exclusively on pitch to illustrate the approach.

3.3.3. Derived representations

Previous research has demonstrated that derived viewpoints yield significant improvements, in terms of both improved prediction performance and fit to human data (Pearce, 2005). For pitch, representations of relative pitch, especially interval (in semitones, see Table **Error! Reference source not found.**), have proved particularly fruitful (Pearce et al., 2010). Another useful pitch representation is octave-equivalent pitch class (MIDI pitch modulo 12, see Table **Error! Reference source not found.**) eliminating octave information. Finally, chromatic scale degree can be computed by making pitch-class relative to a tonal centre such that, for example, 0 is the tonic, 2 the supertonic and so on (see Table 2). Linked viewpoints combining pitch features with rhythmic features have also shown their worth (Pearce, 2005), suggesting that pitch structure and rhythmic structure tend to be correlated.

Table 2. Examples of the melodic representations for the first phrase of Giovanni Maria Nanino's *Artifex mirus* (canto voice). Note that the Krumhansl-Schmuckler key finding algorithm returns the key of A minor for the whole of *Artifex mirus*.



Onset time	0	4	5	7	11	13	15	17	18	19	21
Duration	3	1	2	4	2	2	2	1	1	2	2
Pitch	C#5	C#5	C#5	D5	F5	E5	E5	E5	D5	E5	E5
Pitch class	C#	C#	C#	D	F	E	E	E	D	E	E
MIDI note number	73	73	73	74	77	76	76	76	74	76	76
MIDI number mod 12	1	1	1	2	5	4	4	4	2	4	4
Chromatic Scale degree (relative to A)	4	4	4	5	8	7	7	7	5	7	7
Pitch interval	NA	0	0	+1	+3	-1	0	0	-2	+2	0

It would also be possible to consider simultaneous harmonic intervals (chords) in the polyphonic compositions or implied tonality in monophonic music. However, this could be problematic in a historical context if the representation is founded on assumptions of functional harmony (Rohrmeier, 2011). Such issues could be circumvented if the vertical sonorities are merely described by a descriptive labelling such as Allan Forte's set-theoretical system (Forte, 1973).

Finally, to compare multiple works in terms of their pitch or interval content, it is prudent to attempt to align them to the same approximate tonal centre, so that meaningful comparisons can be carried out. Naturally, concepts of tonality may not apply in the same way to all music corpora, across historical periods, but at the broadest level most musical cultures using discrete pitches rely on hierarchical organisation of pitch classes (Stevens, 2004). For this reason, we prefer to use (chromatic) scale degrees by making the pitch classes relative to a tonal centre. Since the key-signature cannot be used to infer the tonal centre, we adopt a perceptual solution and transpose each work using the Krumhansl-Schmuckler key-finding algorithm (Krumhansl, 1990), which compares the pitch class distribution of the work to the 24 possible key profiles (Krumhansl & Kessler, 1982) and chooses the one with the highest correlation.

3.3.4. Representing polyphony

A limitation of the system described above is that it only applies to melody, where the sequence of notes making up the context for prediction is unambiguous. Although research has extended multiple viewpoint frameworks to homophony especially in voiced music (Conklin, 2002), the problem of representing polyphonic music of arbitrary complexity in an appropriate way for statistical modelling remains unsolved to date. Although this is not problematic for some musical styles (e.g., plainchant), any complete representational system must handle the complexities of true polyphony. In particular, one has to address the question of streaming (Bregman, 1990): one cannot always assume that all voices fuse together into a single stream of harmonic movement. We can identify three broad approaches for applying statistical models to early polyphony:

1. treat each voice independently;
2. assume a single homophonic harmonic movement;
3. identify points at which different streams segregate and integrate perceptually prior to modelling.

The first approach, in which each part is represented individually as a separate monody, could be useful, for example, in making comparisons between different voices. A variant of this approach would be to attempt to identify the melodic line likely to be perceived, either by taking the highest voice or using some automated procedure such as the skyline algorithm (Uitdenbogerd & Zobel, 1998). The second approach involves developing a representation scheme for harmonic movement within the multiple viewpoint framework. Some progress towards this goal has already been made (Conklin, 2002; Whorley et al., 2010). However, neither of these approaches is capable of capturing relevant structural relations between the voices in music such as counterpoint (from the Renaissance onwards). The final approach, therefore, is the most complex, since it effectively involves building a cognitive model of stream segregation prior to analysis. A model capable of combining both horizontal and vertical constraints will be better able to capture voice-leading constraints in counterpoint, for instance. (See Huron, 2001, for an analysis of voice-leading principles in terms of experimentally established perceptual principles.) Nonetheless, it is worth noting as a goal to strive towards. In the meantime, the pertinent question is: to what extent can useful progress be made using the first two options?

3.3.5. The present representation

In the present approach, we focus on melody to demonstrate the basic principles of the approach. We represent melodies in terms of sequences of chromatic scale degrees (relative to a tonic, see Table 2) and sequences of pitch intervals (representing the difference in semitones from the previous pitch in the melody, see Table 2) which have been used in previous work (e.g., Hansen & Pearce, 2014; Pearce et al., 2010). For scale degree representations, pitch-classes are represented relative to the tonal centre indicated by a key-estimation algorithm (Krumhansl, 1990). In computing the tonal centre, if there are several melodic lines in the excerpt, we aggregate the counts across the voices by adding up the distributions across the voices and divide them according to the total number of events. Although both pitch representations (scale degree and pitch interval) are simple, there is much evidence to suggest that interval representations are important in the perception of musical structure (Dowling & Bartlett, 1981; Müllensiefen & Frieler, 2007). We represent intervals across rests (effectively ignoring rests), on the assumption that listeners perceive melodic intervals across rests (although this may not be true for the longest rests in the corpora, the effects should be relatively small).

Scale degree or pitch interval frequencies can be computed in a way that incorporates the transitions between sequential events. When we go beyond the zeroth-order statistics (unigrams) – which are merely the frequency counts of either scale degrees or intervals – higher-order representations reflect the transitions between the tones and intervals (1st-order statistics or bigrams), sequences of three tones/intervals (2nd-order statistics or trigrams), up to sequences of five tones/intervals (4th-order statistics or 5-grams). These higher-order models typically reveal the structural particularities of specific musical styles to a greater extent than the general patterns usually evident in zeroth-order statistics. Moreover, listeners have been shown to be sensitive to such patterns in music and the higher-order statistical information is something that listeners with appropriate stylistic knowledge can utilise in predicting musical continuations (Eerola et al., 2009).

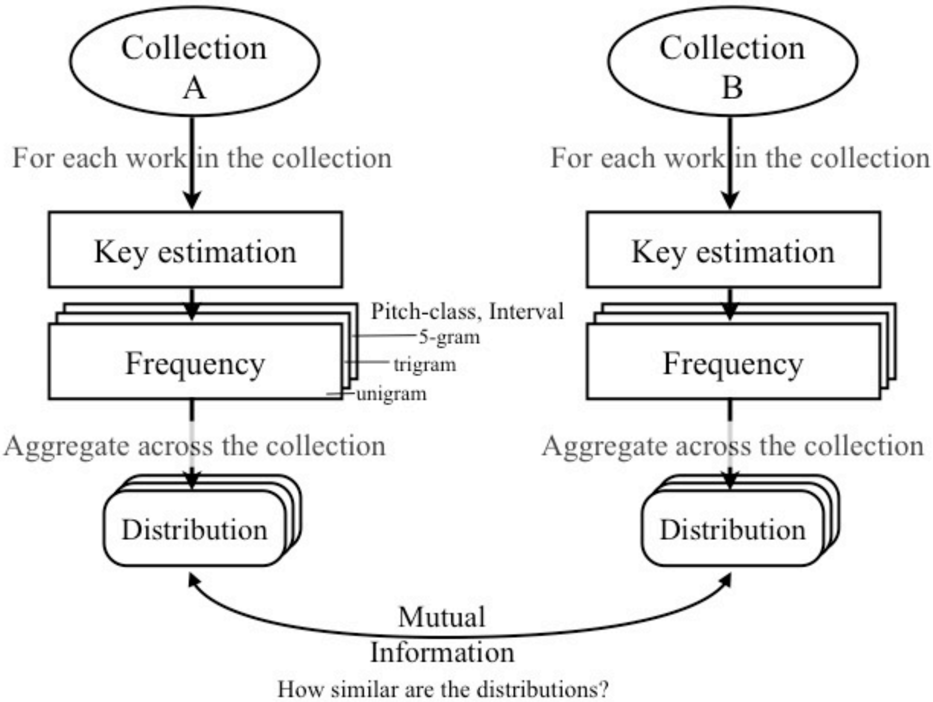


Figure 1. A schematic diagram of the analysis and comparison process.

Higher-order global measures of note and interval transitions might be good candidates for representing the long-term knowledge of a particular style. For instance, the 5-grams for these collections tend to highlight more structurally specific aspects of melodic structures such as the typically descending scalar interval motifs in the *Early*

Renaissance collection (the most prevalent of these being -2 -1 -2 -2 2 and its close alternatives, expressed in semitones) compared to prominent motifs based on note repetitions in popular music (0 0 0 0 0 and 0 0 0 0 2 being the two most common 5-grams in the *pop songs*) or repetitions with descending minor thirds in the *Native American collection* (-3 0 0 0 0 and 0 -3 0 0 0). However, the higher the order of the *n*-gram, the more structurally specific and exclusive the information represented, which can impact negatively on generalisation to music not appearing in its training set.

We can explore the utility of the different representations and specificities of the global measures by comparing their similarities to each other. In order words, we want to assess whether the LTM representations we construct are distinct enough to simulate enculturated listeners from different places at different points in time. Figure **Error! Reference source not found.** shows each of the steps involved in the overall process of comparison (key estimation, computing unigram, trigram and 5-gram frequency counts for scale degrees and intervals, aggregating across the collections and then comparing the resulting distributions). In aggregating across collections we simply summed and normalised the distributions of the individual works of each collection. To effectively compare the different distributions to each other, we need a measure that is especially sensitive to small variations between the distributions. One such measure is *Mutual Information* that captures the mutual dependence of the two distributions (MacKay, 2003):

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{4}$$

Mutual Information ranges from 0 (indicating complete independence) to *min*(*H*(*X*),*H*(*Y*)) (indicating complete redundancy between X and Y). We use a normalised version, *(H(X, Y) – I(X; Y)) / H(X, Y)* that varies between 0 and 1 and can be used as a distance measure. Other related information-theoretic measures exist (e.g., Kullback-Liebler or Jensen-Shannon divergence) which can be examined in future research.

3.4. Selecting the Model Parameters

In many cases, it makes sense to use a combination of short and long-term models both because this improves prediction performance (Pearce, 2005) but also because it makes for a more plausible model of human perception, combining short-term learning of local structure in the current listening episode with schematic implicit knowledge learned over a longer period of experience with music. In general, BOTH+ is the most flexible model and tends to yield the best prediction performance (Pearce, 2005). The LTM+ configuration often achieves comparable performance to BOTH+ but does not allow us to delineate and separately examine the short- and long-term influences on expectations.

In the present work we limit ourselves to simple LTMs trained on different stylistic collections of music. Since we are interested in relationships between collections of music, rather than within individual compositions, the STM is less relevant to our enquiry. Because these collections are static, we prefer the LTM to the LTM+.

We are now ready to train models on the selected corpora to simulate different listeners.

4. Results

For our illustrative examples, we use several LTMs trained on the music of a given period to simulate a listener from the time and location of the stylistic tradition. By comparing models trained on different historical corpora, we can start to examine the model parameters required to simulate adequately an historical listener enculturated in a given musical tradition. We will first look at the specificity of the simulations at the level of entire collections, then examine relationships amongst the musical pieces in the collections, and finally we apply the models to a specific excerpt from the *Coppini collection*. We remind the reader that these examples are intended to illustrate the approach and provide a springboard for further discussion and development rather than provide definitive results.

4.1. How specific are the long-term models?

If we look at the basic representations presented above, a question arises about how much statistical structure is shared or distinct between different corpora? For instance, simple zeroth-order (unigram) models of pitch class and pitch interval in folk music spanning different continents are to a large degree indistinguishable (Huron, 2001). Such broad characterisations are assumed to reflect basic organisation principles that are related to human production and perceptual systems. Small intervals, for instance, are favoured due to constraints of tessitura and vocal range (von Hippel, 2000) and the organisation of frequencies into discrete pitches that are organised hierarchically across the octave is another heuristic facilitating memory (Kessler et al., 1984). For this reason, it is likely that such simple summaries are not particularly good long-term models of stylistically-specific structures in musical traditions. To demonstrate this, Figure **Error! Reference source not found.** displays the interval distributions (unigram) of all 7 collections selected here. This not only illustrates how small intervals dominate but how small the differences are, from casual visual inspection, between many of the distributions (in particular *Essen folk songs*, *Pop songs* and *Schubert & Foster songs* are indistinguishable). The *Coppini collection* and *Native American collection* stand out in this comparison by their frequent note repetitions.

To explore the utility of different representations (basic and derived) and specificity of the higher-order *n*-grams, we calculated the similarity between the *Coppini collection* and the comparative LTMs (aggregated *n*-grams of other collections), using both scale degree and pitch interval representations with a range of model orders (unigram, trigram, 5-gram) according to the process illustrated in Figure 1. We define similarity using mutual information, which reflects the mutual dependence of the two distributions. This exercise, provided in Table 3, underscores the assumption that low-order statistics (unigrams, even trigrams here) are unable to bring out specific differences between the LTMs since most of the collections yield a range of highly similar values to *Coppini collection* (for unigrams based on intervals, all are between .59 and .80).

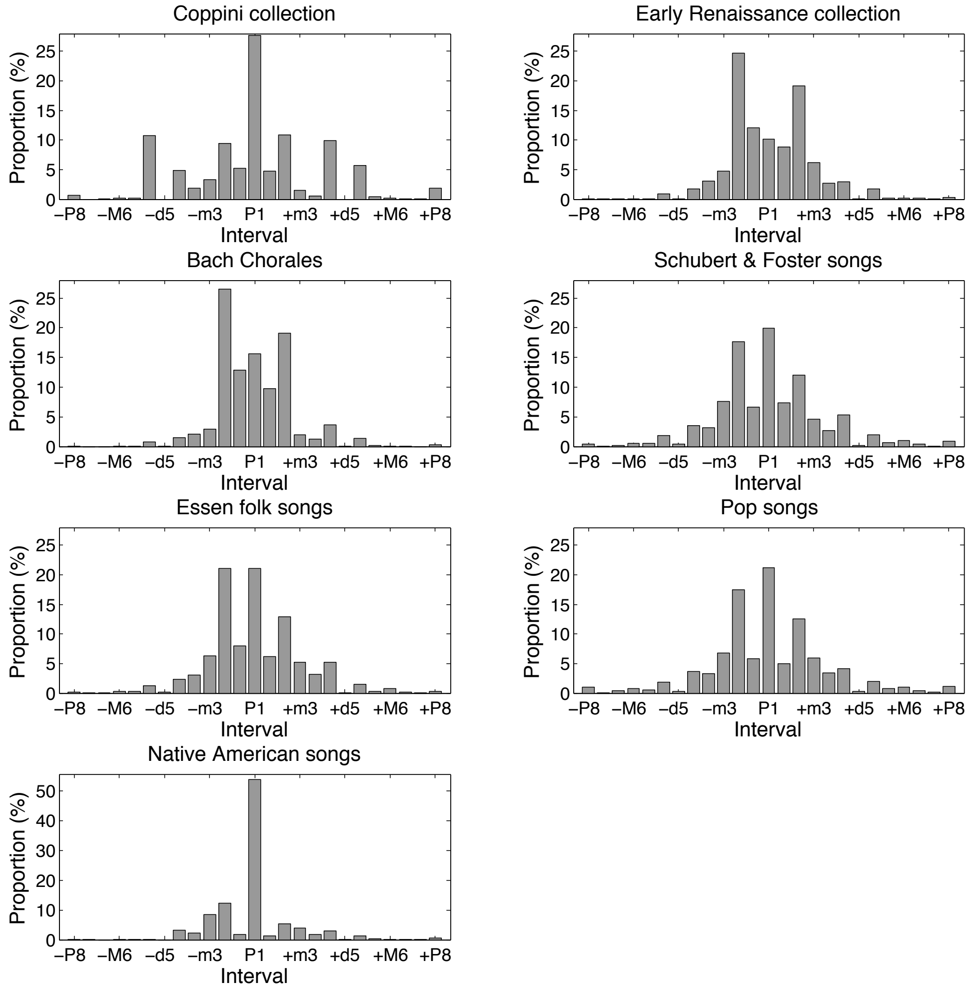


Figure 2. Interval distributions across the collections. Horizontal axis refers to the interval size and direction in semitones (from descending octave, -P8, to unison, P1, at the middle, to the ascending octave, +P8).

In contrast, the higher-order statistics (5-grams) display a rudimentary form of stylistic sensitivity in showing that the *Early Renaissance collection* has the closest relationship to the *Coppini collection* (interval-based 5-grams, 0.80) whereas the other collections show decreasing similarity (from .70 to .59). These results are useful in determining how to choose LTMs that are specific enough to represent musical knowledge specific to a particular style of music. Although these simple models trained on limited numbers of musical examples are far from perfect simulations of the knowledge possessed by an enculturated listener, they do at least allow us to analytically assess the implications of building LTMs from different corpora to simulate listeners with different cultural and historical sensitivities.

Table 3. Similarity of each collection to the Coppini collection.

	Similarity with the Coppini collection					
	Scale degree			Pitch interval		
Collection	unigram	trigram	5-gram	unigram	trigram	5-gram
Early Renaissance collection	0.89	0.72	0.67	0.59	0.73	0.80
Bach Chorales	0.67	0.91	0.58	0.80	0.68	0.70
Schubert & Foster songs	0.55	0.60	0.74	0.75	0.76	0.76
Essen collection	0.61	0.60	0.76	0.73	0.78	0.66
Pop songs	0.66	0.77	0.64	0.72	0.74	0.68
Native American songs	0.82	0.60	0.71	0.60	0.50	0.59

Next we present an analysis at a more detailed level, examining differences between individual compositions rather than collections.

4.2. How consistent are the long-term models?

As well as looking at the collections as a whole, represented as averaged *n*-grams, we are also interested in how individual works relate to other pieces both within and between collections. There are numerous techniques within the field of machine learning for conducting such an analysis, particularly if we want to know which features set the collections apart. Instead of such a discriminative analysis, here we aim to examine the individuality or distinctiveness of compositions from each collection. We will use one of the representations introduced above (trigrams based on pitch intervals, a useful compromise between structural specificity and statistical power) and take a sample of 100 items from each collection (except the *Coppini collection* and *Schubert & Foster songs*, which are present in their entirety). We calculate the similarity between all pairs of items (597 in total) and after converting these into pairwise distances, project them using Multi-Dimensional Scaling (MDS) into a low-dimensional space[viii]. The result is visualised in two-dimensions in Figure **Error! Reference source not found.**.

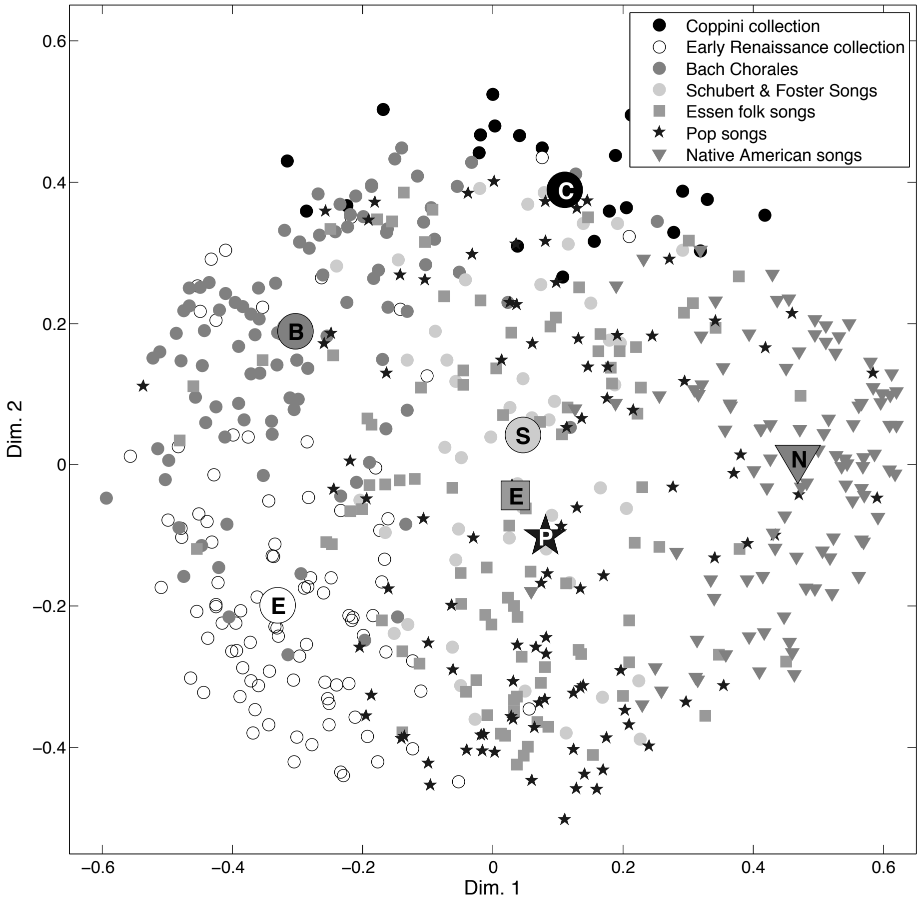


Figure 3. A projection of the similarities between the 100 examples from each collection based on the scale degree trigrams. The axes are simply the first and second dimensions extracted in the MDS projection of the original data (and have arbitrary units). Our simple interpretation of the axes (see text) is that the abscissa (Dim. 1) represents the richness of the interval palette while the ordinate (Dim. 2) represents a continuum from unisons to scalar sequences.

The labels in Figure 3 indicate the central points for each collection in the projected space computed by the MDS, which reveals how *Coppini* and *Early Renaissance collections*, and *Bach Chorales* and *Native American songs* form fairly distinct clusters whereas *Schubert & Foster songs*, *Essen folk songs*, and *Pop songs* are diverse and scattered widely across the projection. Items close together in the projection are more similar in terms of the trigram pitch interval model. Even though this model was not the most discriminant in separating the other collections from the *Coppini collection* using mutual information (Table **Error! Reference source not found.**), interesting relationships between the higher-order interval structures can be interpreted from this visualisation. One simple interpretation is that the vertical axis represents a continuum from unison to scalar sequences: Compositions situated in the upper part of the projection exhibit repeated interval combinations frequently and employ a wide variety of intervals whereas those at the other extreme (lower half) tend to draw from a more constrained set of intervals but less frequently rely on unisons. A simple interpretation of the horizontal axis would be to relate it to the richness of the interval palette: Compositions on the extreme left contain the most varied palette of three interval combinations whereas the examples on the right use a restricted palette of interval chains.

If this interpretation is correct, it is perhaps not surprising that polyphonic works bear a close relationship to one another: the *Early Renaissance collection* and *Bach Chorales* partially overlap, and the items in the *Coppini collection* overlap with *Bach Chorales*. *Native American songs* seem to be most distinct from the other collections. Compositions from the remaining three collections are scattered across the projection, which reminds us that the collections do contain stylistically diverse material (e.g., characteristic differences between *Schubert* and *Foster* songs, the variety of genres represented in *Pop songs*, and the range of temporal and geographic cultures represented in the *Essen collection)*.

Although this visualisation is far from complete (e.g., it relies on a single pitch interval feature, we retain only two dimensions in the projection and so on), it does reflect the fact that the collections are not monolithic entities. Different features, models, similarity measures and projection techniques may result in different projections. Properly and rigorously developed, however, this approach can lead to interesting insights if used in conjunction with historical and stylistic information about the crucial structures within and between compositions. For example, non-traditional authorship attribution could be used to attribute unknown works to the stylistic signatures exhibited by a composer or a geographical group of composers (Dor & Reich, 2011; Jürgensen & Knopke, 2006).

Next we will apply the stylistically close and distant LTM models to note-by-note expectations to investigate the experience of simulated listeners throughout listening to an individual composition.

4.3. Application of two long-term models to selected examples

Next we illustrate how the models representing different LTMs might act as a useful diagnostic tool in explaining what listeners exposed to these styles of music would expect as typical melodic movements. The first example comes from the Coppini collection, bars 1 to 13 from Giovanni Nanino's *Artifex Mirus* (the Canto voice). To contrast two LTMs, one closely associated with this repertoire, another less so, we take the Coppini collection (minus *Artifex Mirus*) as the style-appropriate LTM and the Essen collection as the unrelated one. Mutual information of the 3-grams based on intervals within each 4-note segment of the canto voice versus the respective LTM representation is calculated for both models to derive a prediction for each note. To illustrate the predictions between the LTMs, we have normalised the mutual information across the piece for both LTMs, which renders the segments comparable (as they have the same scale) and easily interpretable (values above 0 suggest that the note is predictable from the LTM while value below 0 suggest that is unpredictable).

Figure 4 illustrates how the first 9 bars of *Artifex Mirus* are more predictable in terms of the Coppini LTM (represented by black bars which tend to have positive values), presumably due to rules dictating species counterpoint in renaissance polyphony (e.g. favour small scale steps, avoid augmented/diminished intervals, skips larger than sixths, see Schubert, 1999). Also, the suspension in the cadential figure in bar 9 with musica ficta and passing notes seems to be particularly typical of the interval patterns in the Coppini LTM in comparison to the Essen LTM.

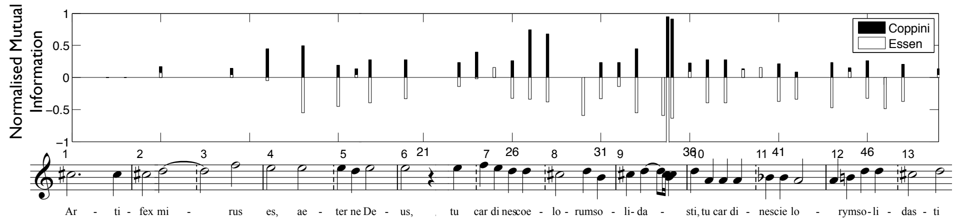


Figure 4. Mutual Information between two long-term memory representations, Coppini (black) and Essen (white) for bars 1-13 from Nanino's *Artifex Mirum* (Canto voice).

This particular cadence pattern is similarly flagged as predictable in other works of the Coppini collection. For instance, Monteverdi's *Pulchræ sunt genæ tuæ* (Quinto voice, bars 93-95), shown in Figure 5, is another example of how the Coppini LTM can pick out the characteristic patterns of the style better than the Essen LTM.

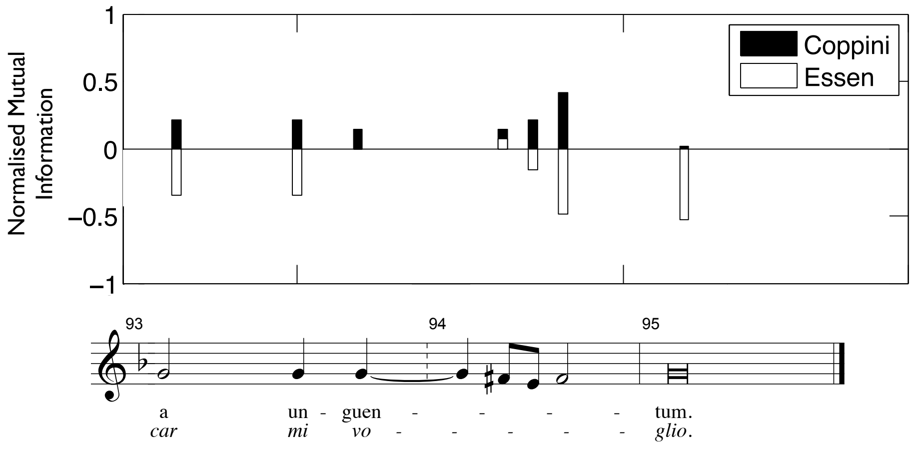


Figure 5. Prediction of two LTM models representing renaissance polyphony (Coppini – black bars) and folk songs (Essen collection – white bars) for Monteverdi's *Pulchræ sunt genæ tuæ* (Quinto voice, bars 93-95).

To illustrate that the LTM models are specific to the styles they represent, let us apply them to an example from the Essen collection. Figure 6 visualises the model output for the tune *Beschattet von der Pappel Weide* (deut2579), where we see the opposite pattern to Figures 4 and 5. The locations containing multiple triadic leaps (the end of bar 1, bar 5 and beginning of the bar 6) are found to be better predicted by the Essen LTM, and again we assume this is due to different stylistic rules and conventions learned by these LTMs from the collections on which they were trained.

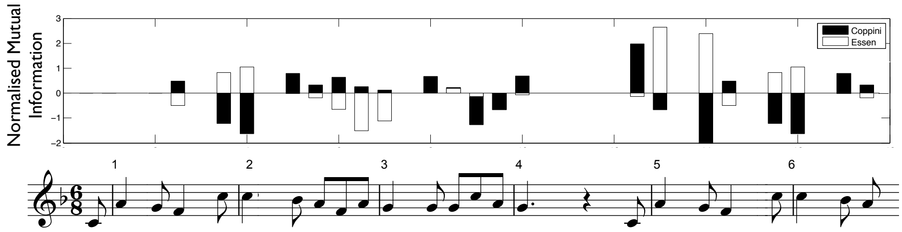


Figure 6. Prediction of two LTM models representing renaissance polyphony (Coppini – black bars) and folk songs (Essen collection – white bars) for a German folktune from the Essen collection (*Beschattet von der Pappel Weide*, bars 1-6, deut2579).

One could also use the models to analytically compare the rules of renaissance composition (avoid augmented/diminished intervals, sevenths, sixths down, large than octave intervals, tritones, etc.), for example, and other stylistic traditions, but the purpose of the present proposal is to use the corpus as a set of examples for implicit learning and cognitive modelling rather than style analysis per se. Nevertheless, corpus-based analysis of this kind might also be used to explore stylistic devices such as *nota cambiata* patterns, or the so-called *Landini cadences* (see Fallows, 2015)

The analysis offered is not without problems, however. We are acutely aware that the comparison of monophonic and polyphonic music is not entirely satisfactory since the voice-leading principles in polyphonic material often lead to large intervals in comparison with monophonic melody that has an accompaniment (pop music). Also, it is somewhat problematic that we are treating the five voices separately and aggregating the results since this probably masks the larger deviations within the individual voices.

A potentially elegant way of handling the rules of counterpoint that constrain pitch in renaissance works is to extend the representation of statistical relationships between, as well as within, voices. Preliminary work by Conklin & Bergeron (2010) has demonstrated that this can be achieved within the multiple framework adopted here. In this multiple viewpoint scheme, probabilities of intervals to other voices at each onset would also be computed and used as an additional distribution in the comparisons. This would have an effect of delineating those harmonic and melodic combinations that are forbidden or rare from those that appear frequently (in particular contexts) in a renaissance corpus without the analyst coding these as fixed rules (Boenn et al., 2012). However, further features would be necessary to account for the principles of voice-leading. For instance, duration and metrical position place constraints on permitted simultaneous or successive intervals in renaissance music. Constructing such multiple viewpoint models is one way to build more accurate cognitive representations of contrapuntal structure than our present simple demonstration with monophonic melodies. Future work should further develop the multiple viewpoint approach for representing such multi-feature contrapuntal constraints.

Finally, it must be kept in mind that we have only used models of long-term memory. The listener will also draw on short-term memory to follow the patterns unfolding and repeating (with variation) within a piece of music. By incorporating short-term memory in the models, most of the frequently repeated sequences of music will become less unpredictable to a computational model (and, presumably, to an enculturated listener). In our simple illustration, for the sake of simplicity, we assume that this short-term memory operates similarly across the models and hence the discrepancies produced by different long-term models would still be evident.

Despite these significant caveats, the purpose of the example is to highlight how the computational models can be used to generate or evaluate hypotheses about how responses to musical passages, as they unfold in time, will vary systematically according to the musical enculturation of the listener.

5. Conclusions

Our goal was to develop a computational framework for developing simulations of historical listeners, allowing us to generate and test predictions about their responses to music. In this approach, a model of musical expectations which has been demonstrated to accurately account for the perception of present-day listeners is applied to the task of understanding the psychological processing of music by listeners at previous points in history. This involves training models on representative corpora of music to simulate a listener enculturated in the musical style represented by the corpus.

We have identified the central methodological decisions that must be made when applying the framework and given a simple illustrative example of how it can be applied, covering three levels of detail. First, we examined differences between simulated long-term models trained on entire collections, finding that higher-order models and pitch interval representations are most discriminative of simulated listeners from different historical periods/locations. The 5-gram pitch interval models exhibited differences between the collections which roughly matched their cultural proximity to the Coppini collection of seventeenth-century madrigals. We then examined similarities between the responses of our simulated listeners to individual compositions, representing those similarity relationships in a 2-d projection. This analysis showed that the *Coppini collection*, *Bach chorales*, *Early Renaissance collection* and *Native American songs* clustered in different parts of the space, whereas the other collections were dispersed more evenly. This analysis, therefore, reveals information about individual compositions within a collection, which are lost when the collection is considered in aggregate. Finally, we examined note-by-note responses of our simulations throughout one madrigal in the *Coppini collection*. This analysis reveals parts of the piece where the simulation of a Pop listener diverges strongly from that of a simulated early Renaissance listener.

A more thorough analysis of particular passages from the Coppini collection could provide further insights into the behaviour of the models. Are the models sensitive to structures thought to be characteristic of a composer's style in historical musicology? How are unexpected melodic events prepared and succeeded in terms of vertical sonorities and can these be represented by the models? For instance, it would be fruitful to add a model of dissonance to the approach, which could combine psychoacoustic aspects of music perception (Vassilakis & Kendall, 2008) and historical understanding of constraints on harmonic intervals in different cultures at different periods (Berger, 2004). Finally, one could ask whether relationships between the lyrics and specific patterns of musical events might influence a listener.

Besides being of interest in its own right, the approach might be useful in addressing issues in historical musicology and music cognition, such as pitch spelling and musica ficta (e.g., Knopke & Jürgensen, 2012; Meredith, 2006), authorship attribution (e.g., van Kranenburg, 2008), stylistic characterisation, development and influence (e.g., Cook, 2007; Volk & de Haas, 2013), and the development of tonal perception (e.g., Huron & Veltman, 2006). We have used relatively simple musical representations (scale degree and pitch interval) in our examples. Other representations of musical structure may be more appropriate; for example, octave equivalence is often not as strong in modal music as in tonal music, suggesting that absolute pitch representations might be useful. In addressing the challenges identified above, therefore, it will be necessary to use more sophisticated representations including:

- different pitch representations;
- both horizontal and vertical constraints in polyphonic music;
- correlations between pitch and rhythmic structure;
- relationships between music and text.

The models we have used to exemplify the approach analyse a collection of music with a very broad brush. However, distinctions between styles often rest on specific rarely occurring details. It would be possible to capture such effects of rareness within and between the music of different historical periods using techniques such as TF/IDF (see e.g., Müllensiefen & Pendzich, 2009, for a musical application). However, it is worth bearing in mind that the goal is to simulate the perception of a listener, which is not necessarily the same as modelling musicological differences between historical corpora.

We have sketched how the framework can be applied at a range of levels: examining whole collections, individual compositions and, finally, specific notes within those compositions. We did not pursue the interpretations of these analyses in great detail, because we believe this will benefit hugely from fruitful collaboration between music history and music cognition. Therefore, our analyses should be taken as illustrative examples of the approach rather than presentations of definitive results.

Future research should extend this sketch in various ways:

1. Can we be more specific - i.e., model the response of particular listeners to particular pieces of music?

2. Can we successfully model complex polyphonic music, requiring more sophisticated representations, including rhythmic patterns and harmonic movement?
3. Is there benefit to using more sophisticated computational models, including short-term models, variable order bounds and multiple viewpoints?
4. Perhaps most importantly, can we use the framework to tackle more complex real-world musicological challenges (conversely, the approach needs historical musicological data to test its predictions about historical listeners)?

If it is possible to address some of these questions, the framework carries great potential for answering fundamental questions about the ways in which early music was perceived and understood by its original audiences.

References

Berger, K. (2004). *Musica ficta: Theories of accidental inflections in vocal polyphony from Marchetto da Padova to Gioseffo Zarlino*. Cambridge: Cambridge University Press.

Bregman, A. S. (1990). *Auditory Scene Analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Boenn, G., Brain, M., De Vos, M. & Ffitch, J., 2012. Computational music theory. In *Musical Metacreation: Papers from the 2012 AIIDE Workshop. Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 27-34. (AAAI Technical Report; WS-12-16)

Butt, J. (2002). *Playing with history: the historical approach to musical performance*. Cambridge: Cambridge University Press.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, *1*(1), 12–29.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*(3), 394–412.

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology*, *52A*(2), 273–302.

Cherla, S., Weyde, T., d'Avila Garcez, A., and Pearce, M. (2013). A distributed model for multiple-viewpoint melodic prediction. In *Proceedings of the International Symposium on Music Information Retrieval* (pp. 15-20). Curitiba, Brazil.

Cleary, J. G. & Teahan, W. J. (1997). Unbounded length contexts for PPM. *The Computer Journal*, *40*(2/3), 67–75.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior. *Psychological Review*, *121*, 33–65.

Conklin, D. (2002). Representation and discovery of vertical patterns in music. In C. Anagnostopoulou, M. Ferrand, & A. Smaill (Eds.), *Proceedings of the Second International Conference of Music and Artificial Intelligence*, volume 2445 of *Lecture Notes in Computer Science* (pp. 32–42). Berlin: Springer.

Conklin, D. and Bergeron, M. (2010). Discovery of contrapuntal patterns. In *Proceedings of the 11th International Society for Music Information Retrieval Conference* (pp. 201-206). Utrecht, Netherlands.

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Cook, N. (2007). Performance analysis and Chopin's mazurkas. *Musicæ Scientiæ*, *11*, 183–207.
Cox, T. F. & Cox, M. A. A. (2001). *Multidimensional scaling*. Monographs on statistics and applied probability. New York, US: Chapman & Hall.
Cuddy, L. L. & Lunny, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgements of continuity. *Perception and Psychophysics*, *57*(4), 451–462.
Curtis, M. E. & Bharucha, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Perception, 26*, 365–375.
Densmore, F. (1909). *Chippewa [i.e., Ojibway] Music*. Bulletin 45 of the Bureau of American Ethnology. Washington, DC: Smithsonian Institution.
Densmore, F. (1918). *Teton Sioux Music*. Bureau of American Ethnology Bulletin 61. Washington, DC: Smithsonian Institution.
Densmore, F. (1926). *Pawnee Music*. Bureau of American Ethnology Bulletin. Washington, DC: Smithsonian Institution.
Dor, O. & Reich, Y. (2011). An evaluation of musical score characteristics for automatic classification of composers. *Computer Music Journal*, *35*(3), 86–97.
Dörffel, A. (Ed.). (1940). *371 vierstimmige Choralgesänge von Johann Sebastian Bach* (4th ed.). Leipzig: Breitkopf & Härtel.
Dowling, W. J. & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, *1*(1), 30–49.
Eerola, T. (2004). *The Dynamics of Musical Expectancy: Cross-cultural and Statistical Approaches to Melodic Expectations*. Doctoral dissertation, Faculty of Humanities, University of Jyväskylä, Finland. Jyväskylä Studies in Humanities, 9.
Eerola, T., Louhivuori, J., & Lebaka, E. (2009). Expectancy in north Sami yoiks revisited: the role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicæ Scientiæ*, *13*(2), 39–70.
Fallows, D. (2015). Landini cadence. In *Grove Music Online*. *Oxford University Press*. Web. 31 Jan. 2015.
<http://www.oxfordmusiconline.com/subscriber/article/grove/music/15943>.
Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387–428.
Forte, A. (1973). *The Structure of Atonal Music*. New Haven: Yale University Press.
Hansen, N. C. & Pearce M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*, 1052.
Hansen, N. C., Vuust, P., & Pearce, M. T. (2013). Predictive processing of musical structure: Effects of genre-specific expertise. *Poster presented at the 35th Annual Conference of the Cognitive Science Society*. Berlin, Germany.
Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*(1), 1–64.
Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
Huron, D. & Veltman, J. (2006). A cognitive approach to medieval mode: Evidence for an historical antecedent to the major/minor system. *Empirical Musicology Review*, *1*, 170–177.
Jacobsen, J. P. (1998). Coppini-samlingen. en udgivelse på internet (www). In T. H. Hansen, B. Marschner, & C. C. Møller (Eds.), *Festskrift til Finn Mathiasen* (pp. 5–10). Århus, Denmark.
Jacobsen, J. P. (2003). Coppini-samlingen fra papir til internet. en beretning om en anderledes kildeudgivelse. (The Coppini collection – from paper to the internet. A different kind of source edition). *Danish Yearbook of Musicology*, *31*, 81–85.
Jürgensen, F. & Knopke, I. (2006). A comparison of automated methods for the analysis of style in fifteenth-century song entabulations. *Musicæ Scientiæ*, *10*(1 suppl), 139–160.

Jürgensen, F., Pearson, D. G., & Knopke, I. (2014-2016). Building an authentic listener. Applying a passive exposure-based training paradigm to detecting differences among compositional styles. *Journal of Interdisciplinary Music Studies*. 8 (1&2): 141-156.

Juslin, P. N. & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–575.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception*, *2*(2), 131–165.

Knopke, I., & Jürgensen, F. (2011). Symbolic data mining in musicology. In T. Li, M. Ogihara, and G. Tzanetakis (Eds.), Music Data Mining (pp. 327-345). San Diego, CA: CRC Press.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Krumhansl, C. L. & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual hymns: Convergence of statistical, behavioural and computational approaches. *Music Perception*, *17*(2), 151–195.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, *76*(1), 13–58.

Leech-Wilkinson, D. (2002). *The Modern Invention of Medieval Music: Scholarship, Ideology, Performance*. Cambridge University Press.

Lerdahl, F. & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception*, *24*(4), 329–366.

Levitin, D. J. & Tirovalas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, *1156*, 211–231.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press.

Macy, L. (2011). Geronimo cavaglieri, the song of songs and female spirituality in Federigo Borromeo's Milan. *Early Music*, *39*(3), 349–358.

Manning, C. D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

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

Margulis, E. (2005). A model of musical expectation. *Music Perception*, *22*, 613–714.

Meredith, D. (2006). The ps13 pitch spelling algorithm. *Journal of New Music Research*, *35*(2), 121–159.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Milne, A. J., Sethares, W. A., Laney, R., & Sharp, D. B. (2011). Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, *5*(1), 1–20.

Müllensiefen, D. & Frieler, C. (2007). Modelling expert's notions of melodic similarity. *Musicæ Scientiæ*, *Discussion Forum 4A*, 183–210.

Müllensiefen, D. & Pendzich, M. (2009). Court decisions on music plagiarism and the predictive value of similarity algorithms. *Musicæ Scientiæ*, *Discussion Forum 4B*, 257–295.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-realisation Model*. Chicago: University of Chicago Press.

Oram, N. & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*(2), 103–118.

Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD thesis, Department of Computing, City University, London, UK.

Pearce, M. T., Conklin, D., & Wiggins, G. A. (2005). Methods for combining statistical models of music. In U. K. Wiil (Ed.), *Computer Music Modelling and Retrieval* (pp. 295–312). Berlin: Springer.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage*, *50*, 302–313.

Reaney, G. (1954). The manuscript Chantilly, musée condé 1047. *Musica Disciplina*, *8*, pp. 59–113.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53.

Rorke, M. A. (1984). Sacred contrafacta of Monteverdi madrigals and cardinal Borromeo's Milan. *Music and letters*, *65*(2), 168–175.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide [computer database]*. Menlo Park, CA: CCARH.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception*, *14*(3), 295–318.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology: General*, *131*(4), 511–537.

Schubert, P. (1999). Modal counterpoint, Renaissance style. Oxford: Oxford University Press.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of Music* (pp. 343–390). New York: Academic Press.

Stevens, C. (2004). Cross-cultural studies of musical pitch and time. *Acoustical science and technology*, *25*(6), 433–438.

Temperley, D. (2007). *Music and Probability*. Cambridge, MA: MIT Press.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*, 885–913.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception*, *16*, 99–118.

Tillmann, B., Poulin-Charronnat, B. & Bigand, E. (2014). The role of expectation in music: From the score to emotions and the brain. *Cognitive Science, 5*, 105-113.

Toiviainen, P. & Krumhansl, C. L. (2003). Measuring and modelling real-time responses to music: The dynamics of tonality induction. *Perception*, *32*(6), 741–766.

Uitdenbogerd, A. L. & Zobel, J. (1998). Manipulation of music for melody matching. In *Proceedings of the sixth ACM international conference on multimedia*, (pp. 235–240). ACM.

van Kranenburg, P. (2008). Assessing disputed attributions for organ fugues in the J.S. Bach (BWV) catalogue. *Computing in Musicology*, *15*, 120–137.

Vassilakis, P. & Kendall, R. A. (2008). Auditory roughness profiles and musical tension/ release patterns in a Bosnian ganga song. *The Journal of the Acoustical Society of America*, *124*, 2448.

Volk, A. & de Haas, W. B. (2013). A corpus-based study on ragtime syncopation. In Britto, A. S., Gouyon, F., & Dixon, S. (Eds.), *Proceedings of the International Symposium on Music Information Retrieval*, (pp. 163–168). Curitiba, Brazil. International Society for Music Information Retrieval.

von Hippel, P. (1998). *42 Ojibway songs in the Humdrum **kern representation: Electronic transcriptions from the Densmore collections [computer database]*. Stanford, CA: Center for Computer Assisted Research in the Humanities.

von Hippel, P. T. (2000). Questioning a melodic archetype: Do listeners use gap-fill to classify melodies? *Music Perception*, *18*(2), 139–153.

Whorley, R., Wiggins, G. A., Rhodes, C. S., & Pearce, M. T. (2010). Development of techniques for the computational modelling of harmony. In Ventura, D., Pease, A., Pérez, R., Ritchie, G., & Veale, T. (Ed.), *Proceedings of the International Conference on Computational Creativity*, Lisbon, Portugal.

---

[i] We thank Tim Carter for pointing out that these are actually contrafacta, Italian madrigals spiritualized by the substitution of Latin texts, which also contain modifications by the editor (usually, but not only, by way of rhythmic alteration to fit the new texts). So it is problematic to suggest that the collection contains madrigals as originally written by the composers listed. For the present purpose this may not matter if the forms used here represent music that people actually heard. As we note above the question of creating such a representative corpus is highly vexed and will require detailed consideration in future research.

[ii] http://kern.ccarh.org

[iii] http://imslp.org/wiki/

[iv] http://kern.ccarh.org/cgi-bin/ksbrowse? l=/users/craig/classical/bach/371chorales

[v] http://kern.ccarh.org/cgi-bin/ksbrowse? l=osu/monophony/{foster,schubert}

[vi] http://kern.ccarh.org/cgi-bin/ksbrowse? l=/essen/europa/deutschl

[vii] http://kern.ccarh.org/cgi-bin/browse?l=/users/craig/songs/densmore

[viii] We use dimensionality reduction (PCA) to reduce the dimensionality of the *n*-grams, the correlation coefficient for assessing the similarity and non-metric Multi-Dimensional Scaling (MDS) (Cox & Cox, 2001).

Biographies

Marcus Pearce was educated in experimental psychology and artificial intelligence at Oxford and Edinburgh and conducted his doctoral and postdoctoral research at City University of London, Goldsmiths, and UCL. He is currently Senior Lecturer in Sound and Music Processing at Queen Mary University of London where he leads the Music Cognition Lab and the EEG Laboratory. He has published widely on computational, psychological and neuroscientific aspects of music perception and cognition, in particular on predictive processing of musical structure and aesthetic experience.

Tuomas Eerola is a professor of music cognition at the Durham University, UK. His main academic background is music cognition that was initially taught within musicology in Finland (MA in 1997), which he supplemented with psychology (BA). His pre-doctoral work involved more emphasis on psychology, during periods of study at the Departments of Psychology at Leicester University (UK) and Cornell University (USA). In 2003, he finished his PhD at University of Jyväskylä (Finland) in musicology (music cognition). The disciplinary frameworks where he mostly publishes are music cognition (computational modelling of melodic processing, similarity, rhythm, and timbre) and music psychology (emotions expressed and induced by music).

Appendix A

Coppini no. 19 p. 1

No. 19. Artifex mirus

Erano i capei d'or

Gio. Maria Nanino

Coppini no. 19 p. 2