journal of interdisciplinary music studies
2014-2016, volume 8, issue 1&2, art. #16081214, pp.121-140

# Open Peer Commentary

## Historically Informed Listening?

*Tim Carter*
University of North Carolina at Chapel Hill
cartert@email.unc.edu

How past listeners responded to music is one of the great unanswered questions facing the Historically Informed Performance movement. HIP used to claim superiority on the grounds of "authenticity" (making music sound now as it sounded then), though it is often accused of being, rather, a commercially opportunistic search for exotic alterity (making music sound different from how it sounds now). HIP would also seem to require some notion of Historically Informed Listening (HIL), but what that means beyond an awareness of contextual referents and stylistic difference remains unclear.

To assume that the processes involved in musical listening have somehow stayed constant over time might seem logical enough *pace* evolutionary developments in neural networks or the like. But any such constancy would seem to undermine HIP as surely as does playing Bach on a grand piano. There are other problems, too. HIP relies on original instruments and historical treatises to reinterpret that elusive element, the music itself. Lacking original ears, and with treatises that tell us remarkably little about how to listen, HIL—making music heard now as it was then—might seem an impossibility.

The present authors adopt the Implication–Realization (IR) model that has gained some currency in music analysis, particularly in Anglo-American (and often, anti-Schenkerian) circles. Its extension into psychologies of musical perception by way of the pleasurable playing on expectations dependent on the predictive capacities of the competent listener has also seemed attractive enough as a heuristic for musical listening. But it is hard to find a coherent basis for it in music-theoretical treatises or the like before 1800, or in many thereafter, save by way of their establishing basic grammatical rules and then allowing certified geniuses to break them.

IR models work most effectively for Western tonality of the common-practice era and, indeed, asserts its hegemony in exclusionary ways. Consider the absurdity of applying them to, say, Messiaen's *Mode de valeurs et d'intensités* (1949), where integral serialism offers the highest degree of predictability though few have the capacity for it. Most non-Western musics also fare poorly under the paradigm given their operational modes, unless we "Westernize" them, with somewhat dangerous overtones in this post-colonial age.

• doi: 10.4407/jims.2016.12.104

The issue is whether the same troublesome hegemonies apply in engaging with historical listening to so-called early music. The conditional probabilities identified by the authors in one particular musical parameter of one specific repertory (with others for comparison) might well be relevant to style analysis or to attribution. There may also be structural issues in play in terms of how certain 3- and 4-grams should reveal the modal functions of the diatessaron and diapente variously distinguished by the position of the *mi–fa* semitone. But as the authors properly note, there is much more going on in this music than can be counted by way of melodic patterning. And as for "How would a seventeenth-century listener have responded to a particular note in a Monteverdi madrigal?" one might perhaps wonder whether this is the right question in the first place.

## Integrating modality into melodic expectation

*Anne-Emmanuelle Ceulemans*
Université catholique de Louvain, Louvain-la-Neuve
Musical Instruments Museum, Brussels
anne-emmanuelle.ceulemans@uclouvain.be

The corpus analysed by M.T. Pearce and T. Eerola, A. Coppini's collection of *contrafacta*, was published in 1607. In order to determine the tonal centre of these musical works, the authors draw among others on Krumhansl and Kessler, who have studied the perception of tonal organisation in a spatial representation of musical keys. This research is based on the twenty-four major and minor keys that are typical for the compositional system of the 18th and 19th century.

However, the relevance of these keys for late Renaissance and early baroque music has often been questioned. Moreover, early 17th century theory is able to offer an alternative explanation for the tonal organisation of contemporary music, the analytical potential of which has been demonstrated by decades of scholarly research on the polyphonic modes (see especially Wiering 2001).

It would be worthwhile to integrate a modal framework into Pearce and Eerola's analytical model, as melodic expectations of historical listeners might have been mode-dependent. From this perspective, Nanino's *Artifex mirus*, supposedly in A major, would be more appropriately qualified as mode 2 transposed an octave upwards. If the final is considered as the tonal centre of the work, this centre would be D rather than A. The frequent use of D as initial and final note in melodic phrases is thus easily interpreted from a modal viewpoint. In A major, on the other hand, it cannot be so easily explained.

Whereas modal theory might explain some melodic expectations of historical listeners, Pearce and Eerola's analyses could in turn shed new light on differences of melodic treatment between the ancient modes.

1. Unlike the tonic and dominant of modern keys, the final and *repercussa*, the second most important note, of the modes are not always a fifth apart. How are these modal poles put forward in vocal polyphony?
2. Modal identification of works ending on A is often problematic. Some of them are evidently centred on D and can be associated with mode 1 or 2, while others rather insist on E, and are thus ascribed to mode 3 or 4. Many compositions ending on A are however ambiguous (see e.g. Mangani and Sabaino 2008, esp. p. 244). Pearce and Eerola's statistical approach might be able to quantify this ambiguity. An example for this within Coppini's collection is A. Gabrieli's *Ne confide*. Melodic and cadential turns at the beginning and the end of the piece point towards D as a melodic centre, but in the middle section E competes with D.

Since the major and minor keys as a theoretical framework become relevant for music of later decades within the 17th century, Pearce and Eerola's analytical methodology might help to chart the evolution of melodic expectations within the gradual emergence harmonic tonality.

## Modelling historical listening: Challenges and benefits through intersections of emotions research, musicology and psychology

*Denis Collins*
The University of Queensland, Australia
denis.collins@uq.edu.au

This study examines a range of issues surrounding the development of computational modelling for simulating historical listeners. Drawing upon a substantial body of existing literature in perceptual studies of music, the authors propose a new model that appears to accurately predict melodic progressions in target repertoire according to perceptual principles. The authors frequently refer to the interdisciplinary potential of their study, especially to the field of historical musicology. They make no claim about providing a comprehensive account of how listeners in past times may have experienced music. As such, the authors concede that they offer a step on the long road towards greater understanding of complex historical and psychological modes of listening. That this road is fraught with potential pitfalls will be clear to readers from the disciplinary perspective of musicology (including myself) when encountering statements scattered throughout the study about possible future musicological enquiries arising from the modelling and results of machine-based simulations. According to the authors, these could include, for instance, questions of authorship attribution or stylistic and structural considerations amongst groups of compositions. These suggestions remain speculative, however, until further research demonstrates otherwise.

The authors note (in section 2.5) the complementary nature of existing research, all drawn from the field of psychology, on different expectations between listeners with diverse musical experiences. However, these findings pertain to present-day listeners, whereas the question of the expectations of historical listeners is generally left unexplored until the authors embark on an exposition of their simulated computational model. Research directions in the history of emotions could provide further leverage for a more nuanced assessment of the complexities of historical experiences of music and how these have changed over time. In particular, the idea of communities of emotional expression has gained traction amongst historians and psychologists and has recently also found its way into musicological discourse. The work of Barbara Rosenwein and William Reddy in particular is being taken up by music researchers for possible directions towards interrogating how specific musical events may be associated with shared emotional responses across listeners. These events could, for instance, include melodic expectation, though much broader parameters of musical activity are also possible. The question of text setting is very relevant here, although the present study only mentions this issue fleetingly in the concluding discussion.

Research on the history of emotions can be considered alongside other more specifically musicological studies on the rich layers of inter-textual meaning within distinct genres (such as the fifteenth-century mass or sixteenth-century madrigal). This would help investigations of historical listening to avoid generalizations about typical listeners in a given period and instead focus on particular contexts or modes of listening that can be probed through an interdisciplinary lens. Overall, however, by showing how a computational model can make reasonably accurate predictions in relation to melodic progressions in a collection of early seventeenth-century madrigals, the present study courageously tackles the very thorny question of historical experiences of music that demands continued attention across several disciplines.

## Modelling potential but questions of psychological plausibility

*Tom Collins*
Lehigh University
tomthecollins@gmail.com

The paper describes a framework – a proof of concept – for studying musical expectancy in historically and geographically diverse groups of listeners. It could be questioned whether such a framework alone contributes much to existing knowledge, without accompanying concrete musicological research questions. The authors acknowledge the preliminary nature of their paper, and that others in the field are using models to engage with real music-historical debates (Knopke & Jürgensen, 2011; Volk & de Haas, 2013). Still I feel it would have been preferable to identify some example research questions and investigate them by applying the framework.

The question that is pursued in the paper, which might be paraphrased as "Does training the model with different data lead to different model outcomes/predictions?", seems too preliminary.

Without a concrete musicological question on which to focus, my attention was drawn to the model itself. Its benefits are said to include (1) incorporation of an experience-driven aspect, and (2) an "ability to combine information from multiple musical features, including tonal and non-tonal effects (all of which are described further below)". One model employed in this paper is trained on chromatic scale degree relative to estimated tonal centre, and another is trained on pitch intervals. There is no combination of "information from multiple musical features", therefore, and so rather than being a tangible benefit, this seems more like the promise of potential future capability.

Also with regards modelling, the treatment of polyphony raises questions (polyphony in the loosest sense of the term, where multiple notes may sound simultaneously). In the proposed approach, a multi-voice work is unfolded so that the melody of the upper-most voice is sequenced first, followed by the next highest voice, and so on. The vertical dependencies between notes are lost, therefore. I do not think it psychologically plausible to assume that this is how any listener perceives a polyphonic work. The authors acknowledge that further work is required to apply their model to polyphonic corpora, and this commentator agrees that models of musical stream segregation/integration may be helpful in this respect.

## Statistical Probabilities and Historical Possibilities

*Jane Ellen Harrison*
Istanbul Technical University
jane.harrison39@gmail.com

Congratulations to Marcus Pearce and Tuomas Eerola for what they have accomplished thus far. The modeling tools and the conceptual framework are flexible and open-ended, making it possible to tailor them to certain listening contexts. Many aspects of their framework combine practicality with finesse, such as the LTM/STM distinction and their possible permutations. As the authors note, these score analysis tools might shed light on the information gaps between musical notation and performance practice in many contexts. Along with the improvements listed in section 5, I suggest working towards representations of timbre, a subject currently of great interests to many music scholars, musicians, and composers.

Expectation norms are often implicated in historical musicological inquiry, necessitating statistical corpus studies in addition to the qualitative analysis of a few salient compositions. Even in the tentative representation that the authors offer here, we see the possibility of operationalizing hypotheses about conventions, innovations, and musico-cultural change in terms of the probability of patterns in different samples. The historical insights in the comparison of melodic expectancy in the six

corpora alone make this article substantially valuable. The richness of the Figure 3 is especially fascinating. A purported musical innovation (*Sturm und Drang* style, Monteverdi's *Orfeo*) and its influence might be operationalized in terms of normative and non-normative expectation schemata tracked chronologically. In instances of purported stylistic conflict, such as Thallberg vs. Liszt (Gooley 2004) or Schoenberg vs. Stravinsky (Gur 2011), do statistically significant expectation patterns support the verbal descriptions of the supposedly incommensurate sides?

*Constructive criticisms and suggestions*

1. Musical scores cannot be conflated with "real life" listening processes, and at times the article falls into this confusion, such as in the title. What have been achieved thus far are possible mental representations of the neutral level (Nattiez 1990) of specific corpora or genres. They can stand as representations of the esthesic level only after being compared against data or textual evidence that refers as directly as possible to the lived experiences of specific listeners.

2. The authors might focus first on modeling listeners who lived in periods after 1800, for which we have a wealth of documentary evidence of individual listening experiences (letters, music journalism) used by musicologists to develop theories of listening specific to certain societies. These data and theories could be used to test the efficacy of the neutral-level expectation models, after which the modeling methodology could be more assuredly applied to phenomena that occurred before 1800, for which we have significantly less textual traces, such as the first listeners of the Coppini collection.

3. Finally, I think the coding procedure of any parameter must be based on culturally-specific experimental and ethnographic research. For the Densmore and pop song collections, I am skeptical that octave equivalence or the Krumhansl key-finding algorithm is culturally relevant. Even when I played through the "Artifex mirus" example I heard the tonal center shifting several times. The mentioned proposal of switching to absolute-pitch coding should be seriously considered. Such ethnocentric biases can make music cognition literature unusable for cultural scholars. Happily, the generally thoughtful research paradigm outlined in this study might make a significant step towards bringing cultural/qualitative and systematic/quantitative music research methods into greater dialogue.

## Probing past musical minds

*David Huron*
School of Music, Ohio State University
huron.1@osu.edu

Implicit learning may turn out to be the single most important tool we have for understanding the musical experiences of past listeners. In their study, Marcus Pearce

and Tuomas Eerola make an initial foray using implicit learning as a lever in in an effort to pry open the minds of historical listeners. The prospects are exciting, even though formidable methodological challenges remain.

The role of implicit learning in music listening has been a major achievement in the field of music cognition. In applying the concept to the experiences of past listeners, an initial question is: "How do we know that people in the past also relied on implicit learning?" Fortunately, animal studies suggest that implicit learning is ubiquitous throughout the animal kingdom: it is unlikely that earlier human generations did not also rely on the same processes.

The next question is: "What music were people exposed to?" The large volume of notated "prestige" music we have inherited from the past is unlikely to be entirely representative. Much of the day-to-day music heard in the pre-sound-reproduction past included unnotated folksongs, drinking songs, and the like. Even among aristocrats, early (formative) musical exposures would have included nursery songs, play songs, and other folk sources for which documentation is spotty. Sampling problems require careful consideration, but with the help of enterprising historical musicologists they are not overwhelming.

With regard to implicit learning, there are many details that remain to be understood. Human minds are not perfect inferential engines. Evidence suggests that some expectations are faulty approximations of objective real-world patterns. For example, experimental observations suggest that the expectation for post-skip pitch reversal is an imperfect approximation of the tendency for melodies to exhibit regression-to-the-mean.

Finally, there are factors other than expectation that are known to play a role in perception and music-induced affect. For example, there are general ethological principles—such as the association of loud+low with aggression, loud+high with alarm, quiet+low with relaxation/sadness, quiet+high with friendliness—that also contribute to perceived or induced affect. Moreover, we can expect unique historically-situated gestures to evoke specific connotations for knowledgeable listeners, such as the various topoi and gestures documented in Topic theory.

How could we ever possibly know what people from the past experienced when listening to music? At first, the task seems Herculean, if not impossible. However, the problem is not unique. On the contrary, the problem is universal and omnipresent: no one has access to more than one mind, so we all make inferences regarding other minds based on a sample of N=1. In both our informal interactions with others, and through formal experimentation, we regularly gain insights into the mind-states of others. These inferences are imperfect and error-prone, but we nevertheless muddle through. Understanding how minds from the past (or minds from other cultures) experience the world is ontologically no different than trying to understand the mind of a person sitting next to you. These are not questions of certainty, but questions of practicality based on the evidence at hand.

We live in an age of happy synchronicity where discoveries in implicit learning are converging with musical big data and techniques in machine learning. The opportunities to address long-standing historical and cultural questions are tantalizing.

## An empirical view of the problem of the 'period ear'.

*David Lewis*
Oxford e-Research Centre, University of Oxford
Department of Computing, Goldsmiths University of London
Institut für Musikwissenschaft, Universität des Saarlandes
D.Lewis@gold.ac.uk

The question of how audience members from a given time and culture perceived musical works on hearing them for the first time is an important area of musicological enquiry. In their article, Pearce and Eerola take some exciting – if preliminary – steps into a rich world. Tools to model the musical experience of a listener past or present – and to address this issue empirically – could be revolutionary, especially if they can harness evidence of musical consumption, such as records of concert performances, and music publishing and sales data.

Such fine-grained, musicologically-strong modelling is currently out of our reach, but Pearce and Eerola show how we might begin. The current approach does have some limitations, which will need to be addressed before its full potential can be realised.

The features modelled are currently fixed-length n-grams of pitch and interval. Multiple voices are treated successively for three of the corpora used, and two of the others have melodies extracted; the remaining two are monophonic to start with. Since writing four or five-part music imposes practical constraints on the composer, especially for the inner voices, the authors' approach carries the risk that the interval patterns of less salient voices will distort the comparisons.

There is good evidence that the parameters modelled here are easily perceived and learned by listeners, and that these colour their expectations, supporting the claims of a perceptual model. It is less well established that this is still the case for pitches and intervals within an inner voice part; the salience of patterns arising from compositional strategy can be hard to assess.

Combining a short and long-term model is a powerful way of modelling the importance of local patterns and of repetition in how we hear music. In future work, this will be useful to help model more complex patterns than the simple repetition of pitch and interval structures discussed in this article, which can be easily thrown off by small elaborations or chromatic alteration that would have very little impact on a listener's sense of musical similarity. Using patterns of repetition (and imitation, which is very hard to treat when parts are taken separately) as a metastructure that can also be statistically modelled could prove a boost to the expressive power of the system discussed in the article.

A cost of making the authors' model richer is that it would become harder to show that it remains a perceptual model; further experiments on whether more complex parameters inform the expectations of listeners are more likely to be dependent on their musical experience.

These comments are more in the nature of a wish list for future work than criticism of what has been proposed or achieved here. By focussing on listeners and their musical experience, the authors are mixing perspectives from both historical and systematic musicology, making an important contribution that can hardly fail to produce further interesting results in the future.

## Nature and tonal structure of corpora

*Rachna Raman and W. Jay Dowling*
The University of Texas, Dallas
rachnaraman@hotmail.com

Pearce and Eerola discuss a groundbreaking model combining music cognition, musicology, and machine learning to characterize how our ancestors understood their music. They demonstrate this model using diverse corpora and simulate the perception of music by historical audiences. Here we comment on three main applications of this model in the fields of music cognition and musicology.

*Nature of the Corpus*

Pearce and Eerola's model represents a first step towards understanding historical audiences' perceptions of music. Future research could embrace more systematically developed corpora that move gradually through time. Collections of scores, such as the Petrucci Music Library, could be mined using a sliding window 50 years wide moving by 10-year intervals from the 16$^{th}$ to the 20$^{th}$ century. Cross-correlations could indicate if the sudden shifts in style reflected in music history textbooks correspond to the data. We could even explore the development of particular composers' styles in relation to predecessors.

*Tonal Hierarchy and Melodic Expectancies*

We could compare the results from the simulations with the responses of contemporary musicians immersed in a particular historical genre, for example, Baroque instrumental music. We could test musicians in period ensembles using Toiviainen and Krumhansl's (2003) concurrent probe-tone technique to examine whether their tonal-hierarchy profiles approximate those of simulations trained on a Baroque corpus versus a more eclectic one. We have used this technique to study perceptions of modulation in both South Indian classical (Carnātic) music (Raman & Dowling, 2016) and Haydn quartets (Dowling, Raman, Ramesh, & Tillmann, 2015), demonstrating the effects of both genre and musical acculturation. Simulations could

generate tonal profiles based on frequency of occurrence and cumulative duration, of pitch classes.

More complex neural network simulations could include polyphonic music, thereby enabling more realistic simulations. Currently, simulations only follow single melodies, which are fine for use with melody-based systems like Carnātic music, which do not involve harmony.

*Pitch classes versus intervals*

Pearce and Eerola refer to an article by one of us (Dowling & Bartlett, 1981) to support their claim that "interval representations are important in the perception of musical structure." Subsequent converging evidence supports a contrasting view: that melodies are perceived and remembered in terms of a melodic-rhythmic contour "hung" on a scale at a particular level (Dowling, 1978; see Dowling, Kwak, & Andrews, 1995, p. 148, for converging evidence). Research (Dowling, 1986; Dowling & Tillmann, 2014) shows that musicians with 5 or more years of training remember melodies this way. That is exactly the population of knowledgeable listeners whose behavior Pearce and Eerola aim to predict.

*Summary*

Here we raise the possibility of extending Pearce and Eerola's model to a systematic exploration of the evolution of musical style, particularly with reference to the use of pitch and time. We also suggest the exploration of how musicians steeped in a period and style hear that music, given their more eclectic knowledge base than that of listeners living in that period. And we emphasize the importance of pitch frameworks (scales) versus isolated pitch intervals in the understanding of musical patterns.

## Assessing the issues in predicting music perception from the past as presented by Marcus T. Pearce and Tuomas Eerola

*Hector Sequera*
Durham University
hector.sequera@durham.ac.uk

The study attempts to address questions in different areas such as music perception (psychology), music in early modernity (musicology), and computer modelling of these two aspects (computer science).

The paper states that the 'differences in expectation between listeners with experience of different musical cultures' are complementary to the 'differences in expectation between listeners with experience of music of different historical periods' (p.10). This may be true and empirical research can be done to verify this hypothesis, i.e. to assess

people familiar with music from the past. However, the next leap in the study is unrelated to these two since the paper recognises 'that we no longer have access to listeners from those periods [the early modern period] for empirical psychological research' (p.12).

Therefore, the creation of 'historical listeners' (e.g. the modelling of audiences from the Renaissance) to sustain the proposed model necessitates some very bold assumptions including: 1) that 'historical listeners' listened and/or perceived music in the same way we do today. This clashes with evidence spanning from Bermudo (1555)[1] to Artusi (1600)[2] suggesting that aspects such as tuning system or dissonances were perceived in very different ways. 2) That 'historical listeners' listened to the same music limited by whatever corpus we come up with (see 'Selecting the Corpus', p.13). This latter point has tremendous implications since most music was probably not notated, and this implies that the influence of everyday music from market places, domestic performances, etc cannot be included in our gathered corpora (this connects with relevant issues of oral traditions and improvisation). The paper briefly acknowledges this (p.14) but does not offer any approach to tackle the issue. 3) However, the most problematic issue is the assumption that the musical score is an accurate representation of what an audience heard in the past.[3] The corpora are gathered from surviving scores that may not represent the culture of the listeners as proposed by the study. This also relates to issues of interpretation: how can we prove that our interpretations of scores in performance are an accurate reproduction of how the music was performed at the time? The latter point links to the idea of 'authenticity', a term that has been addressed by music scholarship since the 1980s (e.g. Dreyfus and Taruskin).[4]

Some of the other limitations of the study are very well articulated e.g. modelling monody vs. polyphony, the limitations of the corpora, modelling of short vs. long term memory, etc. Despite these limitations the methodology remains an intriguing set that could be further refined to work on current issues of music psychology and

---

[1] Bermudo deals with many issues of performance practice such as tuning systems, transpositions, instrument sizes, etc all of which affect the way music is perceived at the time. See for instance Juan Bermudo, *"On Playing the Vihuela" From Declaración De Instrumentos Musicales (Osuna, 1555),* trans. Dawn Espinosa (Lexington, VA: Lute Society of America, 1995). Other relevant ideas on this can be found in Anne Smith, *The Performance of 16th-Century Music: Learning From the Theorists* (Oxford: Oxford University Press, 2011).

[2] The main argument of Monteverdi's treatment of dissonance can be found in Giovanni Maria Artusi, "Artusi, or, of the Imperfections of Modern Music," in *Source Readings in Music History*, ed. Leo Treitler and Oliver Strunk (NY: W. W. Norton & Company, 1998), 526-34. Some of the other issues such as improvised polyphony and ornamentation, tuning, etc are briefly addressed in Claude V. Palisca, "The Artusi-Monteverdi Controversy," in *The New Monteverdi Companion*, ed. Denis Arnold and Nigel Fortune (London: Faber and Faber, 1985), 127-58.

[3] Some authors have devoted entire books to this idea. See for instance Barthold Kuijken, *The Notation is Not the Music: Reflections on Early Music Practice and Performance* (Bloomington: Indiana University Press, 2013).

[4] See for instance Laurence Dreyfus, "Early Music Defended Against Its Devotees: A Theory of Historical Performance in the Twentieth Century," *The Musical Quarterly* 69, no. 3 (1983), 297-322. Taruskin wrote at length about this and his essays are compiled in Richard Taruskin, *Text and Act: Essays on Music and Performance* (NY - Oxford: Oxford University Press, 1995).

ethnomusicology. However, trying to model the past audiences is akin to the performer/musicologist attempting to recreate the music of the past. Taruskin has summarised this best: '…what we call historical performance is the sound of now, not then. It derives its authenticity not from its historical verisimilitude, but from its being for better or worse a true mirror of late-twentieth century taste.'[5]

# Author's Response

## Modelling historical audiences: What can be inferred?

*Tuomas Eerola[1] and Marcus T. Pearce[2]*
[1]Durham University, Durham DH1 3RL
tuomas.eerola@durham.ac.uk
[2]Queen Mary University of London, London E1 4NS
marcus.pearce@qmul.ac.uk

*Introduction*

The target article ("Towards predictive models of music perception in historical audiences", TPM for short) made a bold interdisciplinary proposal and received a varied set of insightful commentaries from scholars in a range of disciplines, for which we are grateful. We have grouped the comments, and our responses, into three categories: first, those that relate to the cognitive modelling of music perception in general (i.e., regardless of time period); second, those that relate to difficulties in making inferences about historical listeners specifically; and third, those that relate to questions of contemporary musical structure and practice.

Before we proceed, we would like to make a few comments about interdisciplinary engagement. Mixing perspectives from separate disciplines such as historical musicology and systematic musicology is challenging due to differences in the underlying motivations, paradigms, terminology and nature of evidence commonly accepted in the disciplines. It is not so much a question of methods, techniques or analyses (these can be learnt and adapted) but whether dialogue between the different disciplines can be established and sustained (e.g., Volk & Honingh, 2012; Clarke, 2009) and whether the questions posed are meaningful across discplinary divides. In TPM, we approached questions that might usually be thought to fall under historical musicology with ideas from cognitive science, psychology and computational modelling. Our original work was not fully interdisciplinary as such, since TPM

[5] Richard Taruskin, "The Modern Sound of Early Music," in *Text and Act: Essays on Music and Performance* (NY - Oxford: Oxford University Press, 1995), 166. These ideas are also articulated in Richard Taruskin, "The Pastness of the Present and the Presence of the Past," in *Text and Act: Essays on Music and Performance* (NY - Oxford: Oxford University Press, 1995), 90-154.

represents mostly the view from systematic musicology, but the open peer commentaries have opened up a constructive discourse concerning the intersection of these disciplines. We hope that our responses below continue this spirit of interdisciplinary engagement.

*Modelling musical minds*

Our central claim from reviewing the existing literature on music perception is that musical expectations reflect a process of implicit learning through which (present-day) listeners acquire cognitive representations of structural relations in the music to which they listen. A second observation from the literature is that computational methods have been developed that allow such structural relations to be learned in an unsupervised manner and these methods have proved quite successful in simulating the expectations of present-day listeners. Our proposal in TPM is that it should be possible to simulate the cognitive representations of historical listeners using the very same computational mechanisms. The proposal rests on two assumptions. The first, is that we have enough relevant information about the music to which historical listeners were exposed. We consider this assumption in detail in a separate section below. The second assumption is that the cognitive process of implicit learning has remained intact over the time period in question (hundreds of years). Huron addresses this assumption, arguing that it would be surprising if earlier generations did not possess the ability, since it appears to be ubiquitous across species. Finally, it is important to note that within the proposed framework, implicit learning of musical structure could, in principle, be simulated by any unsupervised machine learning methods. We have focused on n-gram modelling because it has proved a powerful framework for modelling music perception in present-day listeners. However, future research will no doubt develop more refined models. Whether they can illuminate the musical perception of historical audiences, performers or composers, remains to be determined but many of the considerations discussed here will still apply.

D. Collins and Sequera raise the concern that research on musical expectation that we use to support our approach has been conducted on present-day listeners. To this point, Huron offers the insight that trying to understand the mind of a listener from the past is ontologically no different to trying to understand the mind of a present-day listener, regardless of their physical and cultural proximity. In both cases, understanding subjective experience is a practical process of inference based on the evidence available. The important difference concerns the nature of the evidence: with present-day listeners, we can test hypothesised cognitive mechanisms with empirical studies. For historical listeners, this is not possible but our framework is designed to attempt to make inferences from the data that is available, which is analogous to the way that contemporary music (e.g. pop music) has been used as data for accounting for expectations and tonality in present-day musical listeners (e.g., Temperley & Clercq, 2013).

Might it be possible to conduct empirical inferential experiments that bear on the question of historical music perception? Raman & Dowling suggest comparing the results of model simulations with responses of present-day musicians immersed in a

particular historical genre (e.g., Carnatic music). This is a highly useful suggestion although it is subject to assumptions, first, that the musical tradition itself has not developed over time and, second, that the perception of such listeners is immune from the experience of listening to music that was not available to historical listeners, and third that the emphasis of the culture-specific features is generally constant across cultures, which may not be the case (e.g., Fritz, 2013). Nonetheless, this is an interesting approach that is likely to produce convergent evidence. Another possibility for empirical testing would be to use a cognitive simulation to make predictions about historical transcription errors made by scribes, given evidence that present-day musicians tend to make transcription errors at points where expectations are disconfirmed (Unyk & Carlsen, 1987).

Huron points to evidence that the structural patterns that listeners represent are imperfect approximations of patterns that actually appear in the music they listen to. In other words, implicit statistical learning is subject to representational constraints on learning. More generally there might be other aspects of music perception that do not depend on implicit learning. If they are not dependent on musical experience, and we can assume that they have not changed in response to evolutionary pressures over the time period in question, then these aspects should be comparable between present-day and historical listeners. Therefore, it should be possible to incorporate them into the modelling framework by obeying such fundamental cognitive processing limitations as short-term memory, octave equivalence, categorisation of frequency onto hierarchically organised pitch levels and principles of auditory segregation (Stevens, 2004). Huron notes that music perception (historical or otherwise) might also reflect factors other than expectation, such as general ethological principles (presumably stable over time) and unique historically situated gestures that do not generalise (i.e., they are specific to a given period, culture and setting).

Conversely, there might be aspects of music perception other than expectation that are shaped by musical experience. If so, it should be possible to simulate these aspects using broadly the same framework we have outlined (though perhaps with differently parameterised models). In this respect, Harrison suggests looking at representations of musical timbre. This is an interesting suggestion although timbre perception is less well understood as a psychological phenomenon than pitch and the relevant aspects of instrumental and performance style may not be extant. However, given relevant historical evidence, this topic could well follow the same line of reasoning outlined in TPM where existing notions of how performers implicitly learn timbral relationships (cf. Tillmann & McAdams, 2004) could be applied to historical listeners. D. Collins suggests looking at emotional experience. Again this is an exciting possibility but emotional experience is an even more thorny area of music perception than expectation, so this might open up more questions and problems of interpretation than it solves. On the one hand, expectation is thought to play a role in the emotional experience of music but emotional experience is also thought to reflect individual episodic memories and learned associations with particular musical styles, amongst other mechanisms (see Juslin & Västfjäll, 2008), which are no longer extant for historical listeners. Could they be inferred?

Several commentators make useful observations regarding the *n*-gram modelling framework that we propose. Carter points to the limitations of fixed-order *n*-grams in terms of structural representation. This is quite correct though we note that these limitations can be addressed with variable-order Markov models, capable of combining information from models of different order and sophisticated multiple-viewpoint representational frameworks that allow models to combine information from multiple different musical features at different levels of representational abstraction (e.g. Conklin & Witten, 1995; Pearce, 2005). Raman & Dowling make an excellent suggestion in this respect to focus on contour and scale degree representations, which have been found to play an important role in memory for melody.

Lewis also proposes extending the approach with short-term modelling to simulate perception of repeated motives within musical works (e.g. Conklin & Witten, 1995; Conklin & Anagnostopoulou, 2006; Pearce, 2005). Although the proposed framework accommodates implicit learning of such intra-opus patterns using the short-term model, our initial simulations were limited to the long-term model for elegance of exposition. Future research should investigate this question directly.

Lewis also makes the useful point that while increasing the complexity of the model might offer advantages in terms of capturing musical structure, it also becomes more challenging to argue that it is a perceptual model. Therefore, research simulating historical listeners using the framework outlined in TPM must proceed in tandem with testing model developments against the musical perception of present-day musical listeners. It was partly for this reason that our simulations did not make use of feature combinations (as noted by T. Collins) even though this is possible within the modelling framework (by virtue of the multiple viewpoints representation scheme). Research has simulated present-day perception of music using multiple-viewpoint models (e.g. Hansen & Pearce, 2014) but doing so adds complexity. We thought that our illustrative examples would have greater clarity using single features. Nonetheless future research using the proposed framework should certainly exploit the full power of the multiple viewpoint representation scheme.

*Musical past*

A remark frequently made by the commentators (Huron, Lewis, Sequera, and T. Collins) is that we do not actually know with any degree of certainty what music historical audiences were exposed to. This is a fundamental issue to be addressed. However, it is not particular to the approach proposed in TPM, since it applies to any study of historical musical listeners. The music that has survived in scores and manuscripts probably represents only the tip of the iceberg constituting the music prevalent in each era, and quite likely the musical content remaining in manuscripts and collections is subject to biases of various kinds. This might be considered, on the one hand, an impasse or, alternatively, an interesting stimulus for pragmatic research to assess how much can be inferred with the available evidence. Take, for example, the question of how large a corpus one needs to simulate a listener from a given culture (historical or otherwise). This could be assessed empirically by using models

trained on different-sized corpora to simulate contemporary listeners from a given culture on a range of tasks, including tests of melodic expectation. However, this would not address difficulties relating to non-representative corpora resulting, for example, from the fact that much of the music that was heard in many historical periods has simply not survived. However, historical musicologists may be able to provide indirect indices of such music repertoire (using, for example, records from publishing, sales, concerts, marketplaces, private collections) that could inform model training by weighting the materials according to their assumed prevalence (see e.g., London, 2013) instead of relying on single instances within an existing corpora, as was done in the current implementation. Harrison recommends that the emphasis of the analysis should be placed after 1800 due precisely to the increased presence of such documentary evidence after this date. This would also present a fruitful opportunity for collaboration between historical musicologists and music psychologists although the inferential problem remains that there is still no full account (let alone recordings) of the music heard by 19th Century listeners.

Even if we could understand the ways that past listeners might have perceived music using well-studied cognitive processes (such as implicit learning, expectation and auditory stream segregation), the evidence in terms of musical materials used to train and evaluate the simulations is subject to various interpretations (leaving open the possibility of misinterpretation). Ceulemans, for example, questions the relevance of using key profiles for analysing music that is fundamentally modal. In TPM, we followed a data-driven approach reflecting the way the pitches (and intervals) within the octave are used in any given corpus, rather than imposing a Major or Minor tonal hierarchy as a universal solution. This rests on the assumption that most scale systems contain hierarchies of tones, reflected in the statistical structure of the music, which should be learnable, regardless of whether the music is modal or tonal (Huron & Veltman, 2006). Future research should investigate this assumption by comparison with explicit representations of tonal and modal pitch representations, both of which can be accommodated with multiple viewpoints. Ceulemans also provides an interesting proposal for how TPM might be applied to resolve the question of why the *final* and *repercussa* are not always a fifth apart, or how such ambiguous endings emerge across history in conjunction with the more functional role of harmony.

In a similar way, D. Collins raises doubts about the usefulness of the analysis applications (authorship, style classification, etc.) due to many potential pitfalls in the process. Naturally pitfalls exists but we do think that many interesting research questions will emerge from a consideration of how best to navigate them and that fruitful answers to those questions will only result from interdisciplinary dialogue – a good example is the suggestion by D. Collins of using text setting to identify modes of listening and avoid inappropriate generalisation across those modes.

*Musical context*

Finally, we address issues of contemporary musical practice or stylistic sensitivity. Harrison points out correctly that the representations used in the simulations must be sensitive to cross-cultural differences and Carter notes, specifically, that Implication-

Realisation models (Narmour, 1990; 1992) only apply to melodic expectations in Western tonal styles. The IR model consists of two systems – a bottom-up system consisting of a set of universal rules of melodic implication and a top-down system, which is sensitive to experience and, therefore, potentially variable between musical styles (extra-opus effects) and pieces (intra-opus effects). In fact, the bottom-up principles tend to reflect regularities in actual music (Thompson & Stainton, 1996) and some, such as pitch proximity are apparent both in the music and in the expectations of listeners in non-Western cultures (Carlsen, 1981; Eerola, Louhivuori, & Lebaka, 2009; Huron, 2001; Krumhansl et al., 2000). Therfore, they may in fact reflect universal physical constraints of performance, such as the difficulty of performing large intervals or tessitura constraints (Russo & Cuddy, 1999), which are subsequently acquired by listeners via implicit statistical learning through exposure. Our approach does not use the IR principles but rather takes advantage of such a process of implicit statistical learning through exposure, without making a distinction between top-down or bottom-up effects (everything is, in effect, top-down and dependent on experience). This means that our approach is capable, in principle, of simulating the expectations of listeners from other cultures or points in history as a function of the music to which they were exposed.

T. Collins highlights the fact that vertical constraints exist between voices and Lewis notes, more specifically, that multipart writing places constraints on the inner voices. It seems likely that such vertical constraints are represented and processed in musical listening, though further research with present-day listeners is required to develop an understanding of exactly how. For these reasons, though understandable given the present state of knowledge, our treatment of each voice independently is inadequate as a representation of musical structure. We acknowledge this and see it as a spur to further research on representation of polyphonic structure in music that is amenable to modelling using unsupervised learning methods of the kind we described. As noted in TPM, we believe that cognitive models of stream segregation (Bregman, 1990) are likely to be useful in identifying the parallel streams of notes and chords that listeners identify in listening to polyphonic music.

*Conclusion*

The target article, the commentaries and this response represent a dialogue between several disciplines of music research. We have attempted to outline an approach to understanding historical musical listeners using empirical tools and methods from the sciences. This is very much the beginning of an interdisciplinary research programme and subsequent developments will establish how successful the approach proves to be. It is certain, however, that if it is to be successful then it will require the collaboration of experts in computational musicology, systematic musicology, historical musicology amongst other disciplines. Therefore, we think this an appropriate point to reflect on the nature of the interdisciplinary dialogue that is likely to prove fruitful.

One thing required is a patient dedication to the task, an acceptance that it takes time to break-down disciplinary boundaries so as to allow true collaboration and a

realisation that different motivations and methods can co-exist side by side and even complement each other. This article itself has been the subject of many discussions over a period of about five years. The research started in 2011 with invitations by Richard Parncutt to contribute to an exploratory workshop on "Cognition of Early Polyphony" which took place in March 2012, funded by the European Science Foundation (ESF). We prepared our own separate case studies of how statistical models of music could be applied to Renaissance music and negotiated the challenge of presenting them to other scholars from such fields as historical musicology, ethnomusicology as well as music cognition and neuroscience. The workshop itself represented an exciting mixture of rich, interdisciplinary engagement and challenging interpretation of unfamiliar goals, terminology and methodological approaches.

In our case, this could well have been the end of the story and a return to our own disciplinary pursuits. However, thanks to the encouragement and support of the editorial team (FW, BT), we decided to venture into an attempt to deliver a new perspective on understanding the perception of historical listeners. In fact, TPM already represents a dialogue between our respective approaches since, at the outset, one of us had focussed on theoretical modelling frameworks, and the other on the application of models to Renaissance music. The review process soon brought back the broader frame of reference already at play in the workshop. The shortcomings of TPM were neatly laid out by the reviewers in three primary areas: (a) the sophistication of musical processing; (b) treatment of aspects of music history; and (c) consideration of the musical context. Similar issues have been raised and elaborated upon in the open peer commentaries and we are grateful for the opportunity to respond to these issues, thereby furthering the interdisciplinary dialogue. In our opinion, this iterative process has proved highly insightful and, on many occasions, served as a reminder of how to engage and communicate with music researchers outside our specific discipline. On a personal note, we often find ourselves involved with other scientific disciplines (e.g., psychology, cognitive science, computer science, neuroscience) which share many methods, concepts and terminology. To be involved in a constructive discussion of research questions between the humanities and the sciences has proved overwhelmingly more challenging but also far more rewarding in terms of the knowledge and understanding that can result.

## References

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

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

Clarke, E. (2009). Boundaries, expectations and empirical research: A commentary on Judith Becker's "Crossing Boundaries". *Empirical Musicology Review*, *4*(2), 71–74.

Conklin, D., and Anagnostopoulou, C. (2006). Segmental pattern discovery in music. *INFORMS Journal on Computing*, *18*(3), 285–293.

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

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*, 341-354. doi:10.1037/0033-295X.85.4.341

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception, 3*, 281-296. doi:10.2307/40285338

Dowling, W. J., and Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*, 30-49. doi:10.1037/h0094275

Dowling, W. J., Kwak, S.-Y., and Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57*, 136-149. doi:10.3758/BF03206500

Dowling, W. J., Raman, R., Ramesh, A., and Tillmann, B. (2015, November). *The time course of responses to modulations in classical music*. Poster session presented at the meeting of Psychonomic Society, Chicago, IL.

Dowling, W. J., & Tillmann, B. (2014). Memory improvement while hearing music: Effects of structural continuity on feature binding. *Music Perception, 32*, 11-32. doi:10.1525/mp.2014.32.1.11

Eerola, T., Louhivuori, J., and Lebaka, E. (2009). Expectancy in North Sami yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicae Scientiae*, *13*(2), 39–70.

Fritz, T. (2013). The dock-in model of music culture and cross-cultural perception. *Music Perception: An Interdisciplinary Journal*, *30*(5), 511–516.

Gooley, D. (2004). *The Virtuoso Liszt*. Cambridge: Cambridge University Press.

Gur, G. (2011). The Spectre of the End: Musical Avant-Gardism and the Philosophical Turn. *International Review of the Aesthetics and Sociology of Music*, *42*(2), 267-284.

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*, 1052.

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., and 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.

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

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

London, J. (2013). Building a representative corpus of classical music. *Music Perception*, *31*, 68–90.

Mangani, M. and Sabaino, D. 2008. "Tonal types and modal attributions in late Renaissance polyphony: new observations", *Acta Musicologica* 80/2

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realisation model*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realisation model*. Chicago: University of Chicago Press.

Nattiez, J. (1990). *Music and Discourse: Toward a Semiology of Music*. Translated by Carolyn Abbate. Princeton: Princeton University Press.

Pearce, M. T., and Eerola, T. (in press). Towards predictive models of music perception in historical audiences. *Journal of Interdisciplinary Music Studies*.

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.

Raman, R., and Dowling, W. J. (2016). Real-time probing of modulations in South Indian classical (Carnātic) music by Indian and Western musicians . *Music Perception, 33,* 367-393.

Russo, F. A., & Cuddy, L. L. (1999, March). A common origin for vocal accuracy and melodic expectancy: Vocal constraints. Berlin, Germany: Paper presented at the Joint Meeting of the Acoustical Society of America and the European Acoustics Association.

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

Temperley, D., and Clercq, T. de. (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, *42*(3), 187–204.

Thompson, W. F., & Stainton, M. (1996). Using *humdrum* to analyse melodic structure: An assessment of Narmour's implication-realisation model. *Computing in Musicology*, *12*, 24–33.

Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustic (dis)similarities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(5), 1131–1142.

Toiviainen, P., and Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception, 32*, 741-766. doi:10.1068/p3312

Unyk, A. M., and Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, *7*(1), 3–23.

Volk, A., and Honingh, A. (2012). Mathematical and computational approaches to music: Challenges in an interdisciplinary enterprise. *Journal of Mathematics and Music*, *6*(2), 73–81.

Volk, A., and de Haas, W. Bas. (2013). A corpus-based study on ragtime syncopation. In Proc. ISMIR (pp. 163-168). Curitiba, Brazil.

Wiering, F. (2001). *The Language of the Modes: Studies in the History of Polyphonic Music*, New York: Routledge.