Hooked: A Game for Discovering What Makes Music Catchy

Burgoyne, J.A.; Bountouridis, D.; Van Balen, J.; Honing, H.

Publication date
2013

Document Version
Final published version

Published in
Proceedings of the 14th International Society for Music Information Retrieval Conference

HOOKED: A GAME FOR DISCOVERING WHAT MAKES MUSIC CATCHY

John Ashley Burgoyne* **Dimitrios Bountouridis†** **Jan Van Balen†** **Henkjan Honing***
* Music Cognition Group, University of Amsterdam, the Netherlands
† Department of Information and Computing Sciences, Utrecht University, the Netherlands
{j.a.burgoyne,honing}@uva.nl {d.bountouridis,j.m.h.vanbalen}@uu.nl

ABSTRACT

Although there has been some empirical research on *earworms*, songs that become caught and replayed in one's memory over and over again, there has been surprisingly little empirical research on the more general concept of the musical *hook*, the most salient moment in a piece of music, or the even more general concept of what may make music 'catchy'. Almost by definition, people like catchy music, and thus this question is a natural candidate for approaching with 'gamification'. We present the design of Hooked, a game we are using to study musical catchiness, as well as the theories underlying its design and the results of a pilot study we undertook to check its scientific validity. We found significant differences in time to recall pieces of music across different segments, identified parameters for making recall tasks more or less challenging, and found that players are not as reliable as one might expect at predicting their own recall performance.

1. INTRODUCTION

'Aha! Yes, it's *that* song!' Many music listeners, even casual listeners, have had the pleasant experience of recalling a song to memory after hearing a few seconds of its 'hook'. Likewise, many casual listeners can tell almost immediately upon hearing a new song whether it will be 'catchy'. Despite the prevalence of these musical instincts, musicology (in the broadest sense, encompassing music cognition and MIR) can provide only a limited understanding of why certain pieces music are catchy and what is distinctive about the hooks within these pieces of music. The concepts of the hook and of catchiness are vital to understanding human musical memory, but they also have implications outside of music cognition. Charles Kronengold, a musicologist, has posited that the characteristics of hooks might vary across genres and, *a fortiori*, that different assortments of hook characteristics might constitute a working definition of genre [11]. In MIR, a better understanding of hooks and catchiness would be useful for music recommendation (all else being equal, the catchier of two tunes is probably the better recommendation), measuring musical similarity (as estimating similarity between the hooks of two pieces of music may be closer to human perception than estimating similarity over complete pieces), generating satisfying segmentations of pieces of music (as hooks tend to mark the start of new sections), and to some extent, fingerprinting (as hooks are the fingerprints for the brain's retrieval system).

The boundaries between catchiness, hooks, and some other musical concepts are fuzzy. One related concept that has attracted a certain amount of empirical research is the *earworm*, songs that are so catchy that they become involuntarily stuck in one's mind [3, 7, 19]. Earworms are a much narrower phenomenon than catchiness, too narrow, we believe, for many MIR applications: Few users are looking for playlists comprising nothing but earworms. Another related concept is so-called hit-song science, which aims to predict the popularity of songs based on their musical content [6, 16]. This area of study, in contrast, is broader than our area of inquiry. Although catchiness is certainly correlated with popularity, many popular songs are quite forgettable, and we are most interested in music that remains in listeners' memories for the long term. This level of cognitive information seems to be right for contributing to the widest variety of tasks in MIR [9].

The definition of a hook itself is also fuzzy, and as musicologist Don Traut has observed, 'When we go further and ask not only "What is a hook?", but "What is it about the music that makes this a hook?", the picture gets even more blurry' [17]. From a cognitive point of view, we define a hook to be the most salient, easiest-to-recall fragment of a piece of music [9]; likewise, we define catchiness as long-term musical salience, the degree to which a musical fragment remains memorable after a period of time. By our definitions, every piece of music will have a hook – the catchiest part of the piece, whatever that may be – but some pieces of music clearly have much catchier hooks than others. In principle, a piece of music may also have multiple hooks: two or more fragments of equivalent salience that are nonetheless more salient than all others in the piece. There is agreement in the literature that hooks start at points of considerable structural change, or in other

The Netherlands Organisation for Scientific Research (NWO) funded this study under grant 640.005.004, 'Cognition-Guided Interoperability Between Collections of Musical Heritage (COGITCH)'. In addition to comments from those who tested the prototype, we received helpful suggestions from many colleagues, among them Fleur Bouwer, Aline Honingh, Berit Janssen, Jaap Murre, Johan Oomen, Carlos Vaquero, Remco Veltkamp, Lourens Waldorp, and Frans Wiering. The data are available to download from `http://mcg.uva.nl/`.

words, at points that we in MIR would consider to be the beginnings of new sections for the purposes of a segmentation algorithm [4, 15]. There is more debate about the duration of hooks. While songwriters will often speak of the hook as the entire chorus, in fact, only a few seconds are necessary for most listeners to recall a catchy song to memory; one study has shown that after only 400 ms, listeners can identify familiar music with a significantly greater frequency than one would expect from chance [12].

We have designed an experiment that we believe will help to quantify the effect of catchiness on musical memory. Because we consider catchiness to be long-term rather than short-term salience, this design posed some important challenges. First, we needed to be able to work with well-known recordings of well-known music in order to capture fragments that have in fact remained in participants' memories for potentially long periods of time. Individual listening histories vary widely, however, and thus this constraint also entailed the ability to use quite a large set of musical stimuli, on the order of 1000 or more. Moreover, listening histories vary with respect not only to what music participants have heard before but also to how well they know particular pieces; as such, in order to obtain reliable statistics, we also needed to be able to support a much larger number of participants than a traditional psychological experiment. Next to becoming a serious alternative to a certain class of lab-based experiments, Internet-based experiments can potentially reach a much larger, more varied and intrinsically motivated participant pool, positively influencing the ecological validity of the results [10]. Furthermore, given that most listeners enjoy catchy music, our question seems naturally suited for 'gaming with a purpose', which has already proven successful for certain tasks in MIR and for machine learning in general [1, 13]. By framing the experiment as a game, we believe we will be able to collect enough data about catchiness to support a robust analysis of recall from musical memory and also to open new possibilities for using content-based MIR to predict musical catchiness.

2. DESIGNING HOOKED

Hooked, as we have named the game, comprises three essential tasks: a recognition task, a verification task, and a prediction task. Each of them responds to a scientific need in what we felt was the most entertaining fashion possible. In this way, we hope to be able recruit the largest number of subjects possible without sacrificing scientific quality.

2.1 Recognition Task

The recognition task is the heart of the game. It stems from the idea that the defining aspect of catchiness its effect on long-term memory. In particular, the easier a fragment of music is to recall after a long period of time, the catchier it should be. Thus, a 'drop-the-needle' style quiz, whereby a piece of music starts playing from a point in the middle and players are asked to recognise it, seemed to be appropriate. As noted above, there is a consensus in the theoretical literature that the hook should start at the beginning of a new structural section (possibly including the beginning of the piece itself), and we extended this idea to limit the number of starting points to a statistically tractable subset: Music will always start playing from the beginning of a structural section. Then the amount of time it takes a player to recognise the piece is a proxy for how easy that section is to recall, or in short, how catchy it is.

Figure 1a illustrates the recognition game as implemented in our current iOS prototype. A piece of music starts playing from the start of a structural section, and players have several seconds to decide whether they know it. While players are listening, points are counting down; the faster players recognises the piece, the more points they can win.

2.2 Verification Task

In a controlled laboratory environment, it might be justifiable to trust subjects to be honest in claiming to have recognised a piece of music. In a competitive game environment, it is not. We needed a task to verify that players have truly recognised the music at the moments they claim so. Most music trivia games, e.g., SongPop, [1] would ask players to identify the title, composer, artist, or year of release, but this type of question would cause serious problems for the scientific goals of Hooked. Many listeners may know a piece of music rather well without knowing its exact title or the name of the performing artist; moreover, even for those users who do know such trivia, the extra cognitive load in recalling it in addition to recognising the music itself would have an unpredictable effect on reaction time.

Ideally, the verification task would be strictly musical, but precisely because we expect players to know the musical material fairly well, finding a strictly musical task was challenging. Playing a new fragment of music and asking the player whether it came from the same song, for example, would likely be far too easy to be a reliable test. Using any kind of audio degradation to make the task harder would likely make it too difficult in cases where the player genuinely did know the song. Using MIR tools to extract melodies, beats, or some other feature would bias the general notion of catchiness unduly toward catchiness as limited to what such an MIR tool can extract.

In the end, we were inspired by the idea that once players have fully recalled a piece of music to memory, they should be able to follow along with the song in their heads for some time even after the music stops playing. Moreover, there is evidence that absolute tempo is part of musical memory, although the error distribution is somewhat skewed in favour of overly fast tempi [14]. In Hooked, as soon as players claim to know a song, playback mutes for a few seconds. During the mute, players are asked to imagine mentally or sing along actively for a few seconds (Figure 1b). When the sound returns, half the time the music returns the correct place (i.e., the mute was genuinely only a mute) and half the time the playback is offset by a few seconds (i.e., an invisible DJ 'scratched the record' during the mute). The player must answer whether the music is in the right place. We believe that over mutes of the duration we are considering

[1] http://www.songpop.fm/

(a) Recognition (b) Verification (c) Prediction

Figure 1: Screenshots from the Hooked prototype. (a) The recognition task is the heart of the game: A song starts from the beginning of an internal musical section, chosen at random, and the player must guess the song as quickly as possible. (b) The sound then mutes for a few seconds while players try to follow along in their heads. When the sound comes back, players must verify that the song is playing back from the correct place. (c) Occasionally, players instead must do the reverse: predict which of two sections is catchiest, in order to store bonus rounds for themselves.

for Hooked, players who truly remember a song should be capable of following along well enough to identify whether the music has returned in the correct place. The primary challenge is finding empirical evidence for the optimal mute time: not so short that one can judge the continuation on the basis of common-sense musical knowledge or timbral characteristics (type 2 error) but also not so long that it would interfere with the speed of imagining that might well be faster than in singing (type 1 error).

2.3 Prediction Task

We have argued here that because the notion of catchiness inherently invokes musical memory, a scientific definition of the term must involve ease of recall. The recognition game seeks to quantify listeners' behaviour on this axis. We would also like to know how well this formal definition corresponds to listeners' informal intuitions for what it catchy and what is not. As such, we decided to include periodic rounds of the game where we turn the recognition task on its head and ask players to choose which of two fragments from the same song is catchier. An image of such a round in our prototype appears in Figure 1c.

As a survey question, this task is pleasant enough, but it was a challenge to integrate it meaningfully into the gameplay. One idea we may explore in the future is adding a social element. For example, we might ask players to try to fool online opponents by predicting which will be the *less* catchy members of each pair and sending those predictions to those opponents for a recognition task; we would then award prediction players the inverse of the number of points their opposing recognition player earns. For the moment, however, we wanted a self-standing game with an intrinsic reward for the prediction task. Our solution was bonus rounds. Each time players complete a prediction task, the chosen fragment is saved in a special buffer for each player. Periodically, the recognition task will enter a bonus round for double points, with a guarantee that the fragment selected comes from the special buffer of prediction fragments. Thus, users who spend time to do a thorough job with prediction tasks can potentially earn many extra points.

3. TESTING SCIENTIFIC SOUNDNESS

We developed a prototype of Hooked on iOS and undertook a pilot study to identify the best values of the free parameters in the design (the maximum time allowed for the recognition task, the length of the mute, and length of the offset used for false returns in the verification task) and to ensure that the scientific assumptions underlying the design were correct. We recruited 26 testers from within our academic networks, 18 men and 8 women, between the ages of 20 and 70. Most participants spent about 45 minutes testing, some at home and some in their offices, some on their own iOS devices and some on ours.

3.1 Musical Material

Although we designed Hooked to accommodate a very large corpus of music, our pilot study required a more constrained set of musical material. We chose 32 songs at random from the 2012 edition of a list of the 'greatest songs of all time' from a popular annual radio programme. In order to avoid licensing problems as the scope of the game expands, we used Spotify's iOS libraries to stream all audio and require a Spotify Premium membership to play. [2] The Echo Nest has a partnership with Spotify that includes a convenient web service for applying the Echo Nest Analyzer to tracks in Spotify's catalogue, [3] and we used this service to obtain estimates of the start times of the major structural sections

[2] `http://www.spotify.com/`
[3] `http://developer.echonest.com/`

in each song. For the 9 songs that were ranked highest on the list, we retained all sections but the first and last (which often contain silence); with these songs, we hoped to be able to show that there is indeed significant variation in recognition time across different sections of the same song. For the next 8 highest ranked, we retained a random sample constituting half of the sections, a compromise position. For the remaining 15 songs, we retained only a random pair of sections; with these songs, we hoped primarily to be able to introduce some variety for the participants so that they would have a better sense of how the game would feel with a full-sized corpus. In total, this procedure yielded 160 song sections to use for the recognition task. From among these sections, we selected just one random pair from each of the 32 songs to use for testing the prediction task.

3.2 Method

During a testing session, testers worked through recognition tasks for each of 160 sections in a random order. For the first 80 sections, we asked testers to play as they would in the real world. For the remaining 80 sections, in order to test the limits of the verification task, we asked testers to try to cheat the system by claiming that they recognised every section as soon as possible, ideally before they actually knew the song. We recorded the reaction times and whether the responses were correct for each tester and each section. Throughout a testing session, testers also had a 20 percent chance of being asked to perform a prediction task instead of a recognition task for any given round and a 10 percent chance that a recognition round would be a bonus round.

During test runs, we changed some parameters of the game after every 10 recognition tasks. Overall, there were eight possible configurations of the parameters, which we presented in a random order to each tester during both the first, 'honest' half of the test run and again during the second, 'dishonest' half. Specifically, each of three parameters took one of two distinct values, which we chose based on preliminary testing prior to the pilot. The maximum time allowed for recognition was either 10 s or 15 s; this parameter primarily affects the feel of the gameplay, but it has some scientific consequences in the rare cases where players need more than 10 s to decide whether they recognise a fragment. The mute time was either 2 s or 4 s; this parameter in principle affects the difficulty of the verification task. The offset for false returns in the verification task was either 15 s or -15 s; this parameter likewise affects the difficulty of the verification task. Testers were informed when either the maximum recognition time or mute time changed so that they could comment on their preferences; testers were not informed about changes in the offset time for false returns so as not to give extra information they could have used to cheat the verification task.

4. RESULTS

Due to personal time constraints, not all participants were able to complete the pilot in its entirety: 4 made it less than halfway through and a further 5 made it less than 80

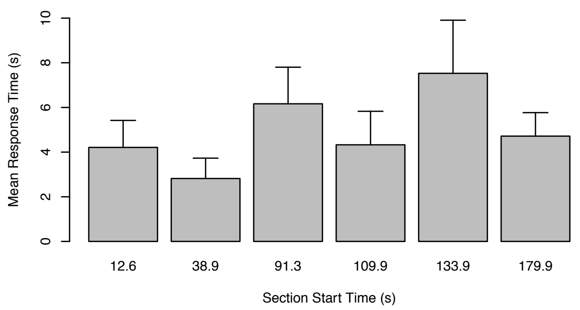


Figure 2: Mean response times from the recognition task on different sections of Adele's 'Rumour Has It'. Error bars reflect standard error. Controlling for multiple comparisons, there are significant differences ($p < .05$) in response times between the bridge (133.9 s) or the verse at 91.3 s, and the initial entry of the vocals (12.6 s) or the pre-chorus at 38.9 s.

percent through. Nonetheless, because we randomised the presentation of sections and parameter settings for each subject, we have no reason to believe that the missing data should exhibit any systematic bias.

For the recognition task, the Box-Cox procedure suggests a log transform on response time, and we assume that response times are log-normally distributed. Regressing thus across all song sections, the average response time for successfully verified claims to know a song is 5.2 s. ANOVA confirms that there are significant differences between the response times for different sections within a song even after accounting for the variation in average response time for different participants: $F(128, 964) = 1.55$, MSE $= 39.06$, $p < .001$. Figure 2 illustrates the variation in response times for Adele's 'Rumour Has It'. [4] After correcting for multiple comparisons with Tukey's test, there are significant differences ($p < .05$) between either of the initial entry of the vocals (12.6 s, 'She, she ain't real') and the pre-chorus at 38.9 s ('Bless your soul, you've got your head in the clouds') and either of the bridge (133.9 s, 'All of these words whispered in my ear') and the second verse (91.3 s, 'Like when we creep out'). The differences in response time are as high as 4 s.

In order to tune the verification task, we needed to determine the best values to use for *maximum recognition time*, the time limit on the recognition task; *mute time*; and the *distractor offset*, the offset to use on the occasions when the sound returns from the mute in the wrong place. More specifically, the distractor offset could be either a *forward offset* of 15 s ahead of where the song should have been playing or a *backward offset* of 15 s before where the song should have been playing. We also needed to ensure that there is a sufficiently large benefit to playing honestly over random guessing. Using the player, maximum recognition time, mute time, the distractor offset, and whether the player was in the 'honest' or 'dishonest' portion of the pilot, we used a stepwise selection procedure on logistic regression models for the probability of answering the validation question correctly. Akaike's Information Criterion (AIC) prefers

[4] `spotify:track:50yHVBbU6M4iIfqBI1bxWx`

Recognition Time	p_1	95% CI	p_2	95% CI
Distractor Offset: −15 s				
10 s	.66	[.59, .73]	.26	[.21, .31]
15 s	.72	[.64, .79]	.19	[.15, .24]
Distractor Offset: +15 s				
10 s	.54	[.47, .61]	.33	[.28, .39]
15 s	.67	[.60, .74]	.33	[.28, .38]

Table 1: Probability of type 1 and type 2 errors for the validation task (i.e., answering the validation question correctly for an unknown song or answering it incorrectly for a known song) under different values of the design parameters. The ideal combination of parameters would minimise both types of error, but some trade-offs will be necessary.

a model including only the player, maximum recognition time, the distractor offset, and the 'honesty' variable with no interactions. A maximum recognition time of 15 s vs. 10 s improved a player's odds of answering the validation question correctly by 31 percent on average (95% CI [5, 62]), a distractor offset of -15 s vs. +15 s improved a player's odds of guessing correctly by 57 percent on average (95% CI [27, 93]), and playing honestly improved a player's odds of guessing correctly by 64 percent on average (95% CI [29, 111]). Table 1 summarises the verification data from the pilot in the more traditional language of type 1 and 2 errors.

In order to analyse the data from the prediction task, we use the fact that after completing a full test run, testers in the pilot had also completed recognition tasks for all fragments offered to them as choices in prediction task. We compared the choices made in prediction tasks to the difference in response times for the same fragments when they appeared in recognition tasks. Although there is a statistically significant relationship ($p = .02$), the effect is small: For each second of difference in response times, the odds of a player choosing the faster-recognised member of a pair during the prediction task increased by only 6 percent (95% CI [1, 12]). Moreover, the variance is quite high. Figure 3a shows the distribution of response-time differences where players chose the first fragment in the prediction task and Figure 3b shows the distribution where they chose the second. Although these distributions are each skewed to the appropriate side, it is clear that players are not necessarily consistent with their behaviour in the recognition task when making predictions.

5. DISCUSSION

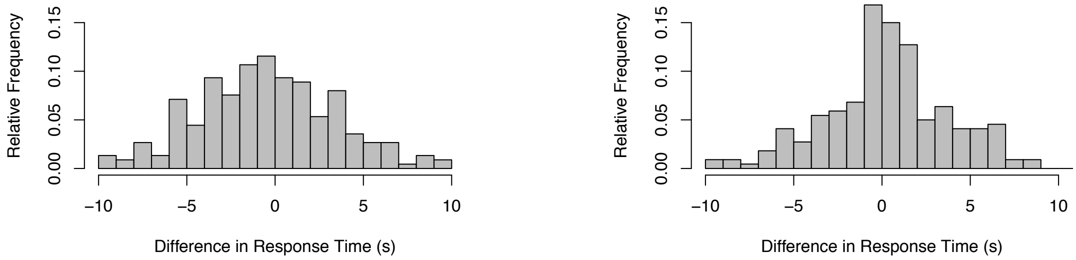
The results of our pilot of the recognition task confirm that different fragments of music, even within the same song, differ measurably in their ability to trigger musical memory. In a context where average response time is just over 5 s, the 4-s effect size is substantial. Moreover, this magnitude of response time sets us comfortably in the realm of musicological theories about hooks: something rather longer than Krumhansl's 400-ms 'plinks' [12] but also rather shorter than a complete refrain chorus, say 5 to 10 s. Historically, MIR has worked rather less with musical fragments of this scale, more often tending to consider audio frames that are shorter even than plinks or attempt to classify complete pieces. Having shown in this pilot study how important these 5-to-10-s fragments are to human musical memory, we would like to suggest that they might be especially profitable when tuning the granularity of algorithms for predicting musical similarity or recommending new music, a claim that is consistent some recent MIR research on segmentation and annotation [2, 5, 8]).

The most important limitation to this result arises from the quality of automatically generated audio segments. If, as musicological theory suggests, hooks are tied to moments of change in the musical texture, any error in the estimation of segment boundaries will propagate throughout the analysis. For a study of this size, it would have been possible to choose the segments by hand, thereby eliminating this source of error, but because our purpose was to test the feasibility of a larger-scale study where it will not be possible to choose segments by hand, we felt it was important to use automatic segmentation for our pilot, too. The analytic techniques available for larger-scale data, most notably the drift-diffusion model [18], will allow us to identify 'lag time' in segments that begin playing a bit too early, but for this study, we have to assume that such lags are noise.

For our verification task, we have arrived at the classical trade-off between type 1 and type 2 errors, perhaps more often encountered in MIR when trying to optimise precision and recall: Because we found no significant interaction between parameter settings and honest play, choosing settings to make the game easier for honest players also will make it easier for cheaters. Conversely, the large benefit to playing honestly – again, a 64 percent improvement in the odds of answering the verification correctly – suggest that we may feel comfortable that players have an incentive to play honestly regardless of the parameter settings and thus can focus on making the game as pleasant for honest players as possible. As such, we intend to allow 15 s for recognition and use the −15-s distractor offset.

We were surprised that the distractor offset had such a strong effect on the players' accuracy, and the idea that distractors from the past are easier to identify as incorrect than distractors from the future is especially intriguing from a cognitive perspective: Is it easier to rewind musical memory than it is to fast-forward? Another possibility, perhaps simpler, is that the forward distractor is more likely to be in the same structural section as the original fragment, whereas because we have chosen our fragments always to start at the beginnings of new sections, the backward distractor will always be in a different one. Assuming that structural sections maintain some degree of timbral consistency, the backward distractor may more often offer timbral clues to the player that something is not right when the sound returns.

The data for the prediction task do not lend strong support our hypothesis that recognition time is a proxy for social intuitions about catchiness. This lack of support is especially surprising given that our concern had originally been more that players would somehow learn to choose fragments that optimised gameplay without touching on



(a) Choosing the First Fragment

(b) Choosing the Second Fragment

Figure 3: Distributions of response-time differences from the recognition task on pairs presented during prediction tasks. There are slight differences in the distribution of differences when the first member of the pair is chosen as opposed to the second, but overall, players do not appear to be consistent with their recognition behaviour when making predictions.

their personal feelings about catchiness; in fact, just the reverse seems to be true. Akin to Williamson and Müllensiefen's work on earworms and Burns's more speculative work, [4, 19], as we roll Hooked out to larger audience and thereby generate a larger database, we plan to find sets audio features that correlate with recognition and prediction performance. The difference between these two sets will help clarify this divergence between listeners' actual long-term musical memories and their expectations of them.

6. REFERENCES

[1] L. van Ahn and L. Dabbish: 'Designing Games with a Purpose', *Communications of the ACM*, Vol. 51, No. 8, pp. 58–67, 2008.

[2] L. Barrington, A. B. Chan, and G. Lanckriet: 'Modeling Music as Dynamic Texture', *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 3, pp. 602–12, 2010.

[3] C. P. Beaman and T. I. Williams: 'Earworms ("Stuck Song Syndrome"): Towards a Natural History of Intrusive Thoughts', *British Journal of Psychology*, Vol. 101, No. 4, pp. 637–53, 2010.

[4] G. Burns: 'A Typology of "Hooks" in Popular Records', *Popular Music*, Vol. 6, No. 1, pp. 1–20, 1987.

[5] E. Coviello, A. B. Chan, and G. Lanckriet: 'Time Series Models for Semantic Music Annotation', *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 19, No. 5, pp. 1343–59, 2011.

[6] R. Dhanaraj and B. Logan: 'Automatic Prediction of Hit Songs', *Proc. 6th ISMIR*, pp. 488–91, London, England, 2005.

[7] A. R. Halpern and J. C. Bartlett: 'The Persistence of Musical Memories: A Descriptive Study of Earworms', *Music Perception*, Vol. 28, No. 4, pp. 425–32, 2011.

[8] P. Hamel, S. Lemieux, Y. Bengio, and D. Eck: 'Temporal Pooling and Multiscale Learning for Automatic Annotation and Ranking of Music Audio', *Proc. 12th ISMIR*, pp. 729–34, Miami, FL, 2011.

[9] H. J. Honing: 'Lure(d) into Listening: The Potential of Cognition-Based Music Information Retrieval', *Empirical Musicology Review*, Vol. 5, No. 4, pp. 121–26, 2010.

[10] H. J. Honing and O. Ladinig: 'The Potential of the Internet for Music Perception Research: A Comment on Lab-Based Versus Web-Based Studies', *Empirical Musicology Review*, Vol. 3, No. 1, pp. 4–7, 2008.

[11] C. Kronengold: 'Accidents, Hooks and Theory', *Popular Music*, Vol. 24, No. 3, pp. 381–97, 2005.

[12] C. L. Krumhansl: 'Plink: "Thin Slices" of Music', *Music Perception*, Vol. 27, No. 5, pp. 337–54, 2010.

[13] E. Law, K. West, M. Mandel, M. Bay, and J. S. Downie: 'Evaluation of Algorithms Using Games: The Case of Music Tagging', *Proc. 11th ISMIR*, Utrecht, the Netherlands, 2010.

[14] D. J. Levitin and P. R. Cook: 'Memory for Musical Tempo: Additional Evidence That Auditory Memory Is Absolute', *Perception and Psychophysics*, Vol. 58, No. 6, pp. 927–35, 1996.

[15] P. Mercer-Taylor: 'Two-and-a-Half Centuries in the Life of a Hook', *Popular Music and Society*, Vol. 23, No. 2, pp. 1–15, 1999.

[16] F. Pachet: 'Hit Song Science', *Music Data Mining*, T. Li, M. Ogihara, and G. Tzanetakis, Eds., pp. 305–26, Chapman & Hall/CRC, Boca Raton, FL, 2012.

[17] D. Traut: '"Simply Irresistible": Recurring Accent Patterns as Hooks in Mainstream 1980s Music', *Popular Music*, Vol. 24, No. 1, pp. 57–77, 2005.

[18] J. Vandekerckhove and F. Tuerlinckx: 'Fitting the Ratcliff Diffusion Model to Experimental Data', *Psychonomic Bulletin and Review*, Vol. 14, No. 6, pp. 1011–26, 2007.

[19] V. J. Williamson and D. Müllensiefen: 'Earworms from Three Angles: Situational Antecedents, Personality Predisposition, and the Quest for a Musical Formula', *Proc. 12th ICMPC*, pp. 1124–32, Thessaloniki, Greece, 2012.