

Dissecting an earworm: Melodic features and song popularity predict involuntary musical
imagery

Kelly Jakubowski, Sebastian Finkel, Lauren Stewart, & Daniel Müllensiefen

Author Accepted Manuscript

Abstract

Involuntary musical imagery (INMI or “earworms”)— the spontaneous recall and repeating of a tune in one’s mind—can be attributed to a wide range of triggers, including memory associations and recent musical exposure. The present study examined whether a song’s popularity and melodic features might also help to explain whether it becomes INMI, using a dataset of tunes that were named as INMI by 3,000 survey participants. It was found that songs that had achieved greater success and more recent runs in the UK Music Charts were reported more frequently as INMI. A set of 100 of these frequently-named INMI tunes were then matched to 100 tunes never named as INMI by the survey participants, in terms of popularity and song style. These two groups of tunes were compared using 83 statistical summary and corpus-based melodic features and automated classification techniques. INMI tunes were found to have more common global melodic contours and less common average gradients between melodic turning points than non-INMI tunes, in relation to a large pop music corpus. INMI tunes also displayed faster average tempi than non-INMI tunes. Results are discussed in relation to literature on INMI, musical memory, and melodic “catchiness”.

Keywords: involuntary musical imagery, earworms, melodic memory, automatic music analysis, involuntary memory

Why do certain songs always seem to get stuck in our heads? Involuntary musical imagery (INMI, also known as “earworms”) is the experience of a tune being spontaneously recalled and repeated within the mind. A growing body of literature has described the phenomenology of the INMI experience (Brown, 2006; Williamson & Jilka, 2013), explored the circumstances under which INMI is likely to occur (Floridou & Müllensiefen, 2015, Hemming, 2009; Liikkanen, 2012a; Williamson, Jilka, Fry, Finkel, Müllensiefen, & Stewart, 2012), and investigated traits that predispose an individual toward experiencing INMI (Beaman & Williams, 2013; Beaty, Burgin, Nusbaum, Kwapil, Hodges, & Silvia, 2013; Floridou, Williamson, & Müllensiefen, 2012; Müllensiefen, Fry, Jones, Jilka, Stewart, & Williamson, 2014). In general, it has been found that INMI is a fairly common, everyday experience and many different situational factors can trigger many different types of music to become INMI (Beaman & Williams, 2010; Halpern & Bartlett, 2011; Hyman et al., 2013; Liikkanen, 2012a; Williamson et al., 2012). However, the initial question posed in this paper of why certain songs might get stuck in our heads over other songs is still not well understood. The reason this question is so difficult to answer may reside with the fact that the likelihood of a tune becoming INMI is potentially influenced by a wide array of both intra-musical (e.g., musical features and lyrics of a song) and extra-musical factors (e.g., radio play, context in which it appears as INMI, previous personal associations with a song, and the individual cognitive availability of a song). The present research examines some of these previously unaddressed factors by examining the musical features of and popularity (e.g., chart position, recency of being featured in the charts) of songs frequently reported as INMI.

Related Previous Research on INMI

Several researchers have examined extra-musical features that increase the likelihood that a song will become INMI. Lab-based studies have found that the song that has been

heard aloud most recently is more likely to become INMI than a song heard less recently (Hyman et al., 2013; Liikkanen, 2012b), and recent exposure to a tune is generally the most frequently reported trigger of INMI experiences in diary and questionnaire studies (Bailes, 2015; Floridou & Müllensiefen, 2015; Hemming, 2009; Jakubowski, Farrugia, Halpern, Sankarpani, & Stewart, 2015; Williamson et al., 2012). Familiarity can also increase the likelihood that a song will become INMI. Byron and Fowles (2013) found that participants who were exposed to a previously unfamiliar song six times were more likely to experience that song as INMI than participants who had only heard the song twice. It is also generally uncommon to experience completely novel music as INMI, although a handful of reports of self-composed music have been found in previous work (Beaman & Williams, 2010; Beaty et al., 2013).

In terms of the features of a melody itself that increase its INMI propensity, a pilot study first presented by Finkel, Jilka, Williamson, Stewart, and Müllensiefen (2010) and further developed by Williamson and Müllensiefen (2012) represents the first empirical investigation in this realm. In this study, 29 songs were collated that had been frequently or recently experienced as INMI by more than one participant in an online survey. Then, 29 non-INMI tunes (songs that had never been named as INMI in the online survey) that were similar in popularity and style to the 29 INMI tunes (based on Gower's similarity coefficient; Gower, 1971) were compared to the INMI tunes in terms of melodic features. Statistical melodic summary features of all 58 songs were computed using the melody analysis software FANTASTIC (Feature ANalysis Technology Accessing STatistics (In a Corpus); Müllensiefen, 2009) and a binary logistic regression was used to predict INMI versus non-INMI tunes based on these features. The results of this analysis indicated that INMI tunes generally contained notes with longer durations and smaller pitch intervals than non-INMI tunes. Williamson and Müllensiefen (2012) suggest that these two features might make songs

easier to sing along with, which relates to another result they reported—specifically, that people who sing more often also report more frequent and longer INMI. The present study will build on the initial findings of Finkel et al. (2010) and Williamson and Müllensiefen (2012) and extend this work by using 1) a larger sample of participants (N=3,000) and songs (200 songs) than Finkel et al. (2010) and Williamson and Müllensiefen (2012) (both studies used N=1,014 and 58 songs), 2) a larger set of melodic features (including features based on statistics of a large corpus of music), and 3) more powerful statistical modelling techniques for both matching of the INMI tunes to non-INMI tunes and classifying INMI versus non-INMI tunes based on their melodic features.

Although the present work is only the second study to examine the INMI phenomenon from a computational, melodic feature-based perspective, this type of approach has been employed successfully by various other researchers in order to explain perception or behaviour in a variety of music-related tasks. For instance, Eerola and colleagues have used melodic feature-based approaches to explain cross-cultural similarity ratings (Eerola, Järvinen, Louhivuori & Toiviainen, 2001) and complexity ratings for melodies (Eerola, Himberg, Toiviainen & Louhivuori, 2006). The following sections will review three specific areas in which feature-based approaches have been used to explain aspects of melodic memory and musical composition, which bear some inherent similarities to the present work on INMI.

Research on Musical Catchiness

Some previous research has addressed the concepts of musical “catchiness” and song “hooks”. Burgoyne et al. (2013) offer a definition of melodic catchiness from a cognitive science perspective as “long-term musical salience, the degree to which a musical fragment remains memorable after a period of time” (p. 1) and a definition of a song hook as “the

most salient, easiest-to-recall fragment of a piece of music” (p. 1). These concepts are not entirely analogous to the INMI experience, which is set apart particularly by its involuntary recall and repetitive nature. However, various parallels may be inherent; for instance, the section of a tune that is recalled most easily as a hook might also be the section that most easily comes to mind when *involuntarily* retrieved from memory.

A variety of popular music books have provided advice from successful musicians based on their own anecdotal experiences of what rhythmic, melodic, and lyrical features contribute to the composition of a good song hook (e.g., Bradford, 2005; Leikin, 2008; Perricone, 2000). One of the first musicological investigations of hooks was conducted by Burns (1987), who compiled detailed qualitative descriptions of how hooks might be constructed using rhythmic, melodic, lyrical, timbral, temporal, dynamic, and recording-based features of a tune. A more recent, large-scale empirical investigation of catchy tunes was distributed in the form of an Internet-based game called “Hooked”, in which participants were asked to judge whether they recognized different sections of songs as quickly as possible. Results indicated that different sections even within the same song differed significantly in the amount of time required to recognize them, thus suggesting some sections serve as better hooks than others (Burgoyne et al., 2013). Additionally, the same research team has examined audio and symbolic musical features of their song stimuli and revealed a number of features related to melodic repetitiveness, melodic “conventionality” in comparison to a corpus of pop music, and prominence of the vocal line as predictors of musical catchiness (Van Balen, Burgoyne, Bountouridis, Müllensiefen, & Veltkamp, 2015).

Research on Musical Features of Song Memorability

Another related area of research has examined the melodic features that enhance recognition or recall of tunes from memory. Müllensiefen and Halpern (2014) conducted a

study in which participants heard novel melodies in an encoding phase and were then assessed on both explicit and implicit memory for these melodies in a subsequent recognition task. This study used the same feature extraction software that will be used in the present research (FANTASTIC; Müllensiefen, 2009). A relevant feature of Müllensiefen and Halpern's study to the present work is that it made use of both first- and second-order melodic features. First-order features are features that are calculated based on the intrinsic content of a melody itself, such as the average note duration, average interval size, or pitch range of the melody. Second-order features, also called corpus-based features, are features that compare a melody to a larger collection or corpus of melodies (generally comprised of music from the same genre or style as the melodies that are being analysed, such as pop songs or folk songs). For instance, one example of a second-order feature might measure to what degree the average interval size within a particular melody is common or uncommon with respect to the distribution average interval sizes within a large corpus of comparable melodies. The use of second-order features allows one to determine whether particular features of a melody are highly common or highly distinctive in comparison to a corpus of music that is intended to be representative of the genre from which the melody is taken.

Müllensiefen and Halpern (2014) conducted a number of analyses using partial least squares regression and found somewhat different patterns of results for predicting explicit and implicit memory for tunes. Explicit memory was enhanced for tunes that included melodic motives that were rare in terms of their occurrence in the corpus and that repeated all motives frequently. In terms of implicit memory, the usage of unique motives in comparison to the corpus was also important, similar to the findings on explicit memory. However *less* repetition of motives, a smaller average interval size, simple contour, and complex rhythms were also important to implicit memory recognition.

Although this study is relevant to the present research, several differences are inherent. Müllensiefen and Halpern's work tested whether certain features of a melody can increase memorability for previously unfamiliar tunes that had only been heard *once* before, in terms of both explicit and implicit memory. In the case of INMI, however, tunes that are often *highly familiar* to participants (and have been heard aloud many times before) are retrieved in a *spontaneous* fashion from memory. Therefore, although it is plausible that some of these melodic features related to explicit and implicit memory for previously unfamiliar music might be implicated in INMI, it is also likely that other features might serve to enhance the spontaneous recall of well-known tunes and looping nature of the INMI experience.

Other studies have investigated the musical features that contribute to memory for melodies through the use of paradigms that seek to identify the point at which familiar songs are identified. Schulkind, Posner, and Rubin (2003) conducted such a study in which familiar songs were played to participants on a note-by-note basis. The positions in a song in which participants were most likely to identify the song correctly included notes located at phrase boundaries, notes that completed alternating sequences of rising and falling pitches, and metrically accented notes. Using a similar paradigm, Bailes (2010) explained around 85% of the variance in her participants' data with second-order features that measured timing distinctiveness and pitch distinctiveness in comparison to a large corpus of Western melodies. While different in their primary research question and experimental paradigm to the present work, the results of these studies nonetheless indicate that assessing memory for melodies based on structural and melodic features can be useful in modelling aspects of music cognition and provide impetus for conducting similar research in the domain of *involuntarily* retrieved musical memories.

A final relevant body of literature has investigated the commercial success of songs, that is, whether certain musical features of a song predispose it toward becoming a “hit.” This literature is sometimes referred to as “Hit Song Science.” One common approach in this research area has been to analyse the acoustic features from recordings of songs in an attempt to predict hits versus non-hits based on these features (Dhanaraj & Logan, 2005; Ni et al., 2011). However, the approach of predicting songs based solely on acoustic features has received some criticism, due to the generally low prediction accuracy rates that have been reported (Pachet & Roy, 2008).

An alternative approach that has been employed is to investigate features of the compositional structure of hit tunes. Kopiez and Müllensiefen (2011) conducted a first exploration into this area by attempting to predict the commercial success of cover versions of songs from the Beatles’ album “Revolver.” They were able to achieve a perfect (100%) classification accuracy using a logistic regression model with just two melodic features as predictors—pitch range and pitch entropy as implemented in the software toolbox FANTASTIC (Müllensiefen, 2009)—thereby indicating as a proof-of concept that compositional features can be useful in predicting hit song potential. In their specific dataset, the combination of a relatively large pitch range with relatively low pitch entropy (measured by Shannon entropy and based on the number of different pitches used among all notes in the melody) was associated with the commercial success of cover versions of a tune. However, as the sample of songs used in this study (only 14 songs all composed by the same band) is very specific, it is unlikely that such a simple classifier would be able to cope with the wide diversity of styles and artists represented across all of the “popular music” genres. A subsequent study by Frieler, Jakubowski, and Müllensiefen (2015) investigated the contribution of compositional features to the commercial success of a larger and more diverse sample of 266 pop songs. The study used a wide range of first-order melodic features to

predict hits versus non-hits. The three most predictive variables for hit songs all related to the interval content of the melodies. However, the classifier used in this work only achieved a classification accuracy rate of 52.6%. This finding suggests that extra-musical factors (such as artist popularity) and audio features (such as timbre) may play a large role in the commercial success of pop music, but also leaves open the question as to whether second-order, corpus-based features might help to further capture the unexplained variance in the data. Hence, the present research employed a similar approach for predicting INMI tunes, but included second-order features in addition to simple first-order and summary features.

Aims of the Research

The main aims of the present work were to collate a large number of frequently reported INMI tunes from an online questionnaire and use powerful statistical modelling techniques to examine features of their melodic structure. A preliminary investigation also explored the extent to which the number of times a song was named as INMI could be explained by the song's popularity and recency, measured using data from the UK Music Charts. Based on the findings from this preliminary analysis, the second part of the research examined the extent to which the propensity of a song to become INMI could be predicted by melodic features of the song, while controlling for relevant popularity- and recency-related variables.

Method

Participants

For the present project, responses from 3,000 participants who had completed a pre-existing online questionnaire on INMI experiences were compiled ("The Earwormery"; Williamson et al., 2012). These participants ranged in age from 12 to 81 years ($M = 35.9$, SD

= 13.1); 1,338 participants were male, 1,644 were female, and 18 did not provide gender information. The Earwormery questionnaire has been completed by 5,989 participants in total to date. The decision to include only the responses of 3,000 of these participants in the present study was a purely pragmatic one, as manual verification of each participant's responses to two open text questions was needed to ensure that songs names were spelled correctly and that each response was a genuine, existing song that had been performed by the artist listed by the participant.

Ethics Statement

The present study and the online "Earwormery" questionnaire were approved by the Ethics Committee of Goldsmiths, University of London.

Materials and Data Preparation

The "Earwormery" questionnaire contained questions about the features and phenomenology of participants' INMI experiences, such as how often they experienced INMI, whether they found INMI disturbing or distracting, and their reactions and attitudes toward INMI (see Williamson et al., 2012, 2014 and Williamson & Müllensiefen, 2012 for additional publications related to this questionnaire). The survey questions relevant to the current project were two open-ended questions: one that asked for the name, artist, and section of the tune (e.g., chorus, verse) experienced as the participant's most recent INMI tune, and one that asked for the name, artist, and section of the participant's most frequent INMI tune.¹

The first step in preparing the data for analysis involved compiling participant responses to the two relevant survey questions. The analysis in the present project did not

¹ Within the actual questionnaire, the term *earworm* was used in all instructions and questions directed to participants, rather than *INMI*, as *earworm* was deemed a more familiar and colloquial term.

distinguish between tunes listed as a “most frequent” and a “most recent” INMI tune, as the aim was to include as many different pieces of music in one sample as possible. Songs compiled within the dataset were also limited to only “popular music” genres (e.g., pop, rock, rap, rhythm & blues, etc.), while excluding such music types as classical, children’s songs, and TV jingles. This is due to the fact that the project utilized popularity and recency variables for each of the songs that were measured in terms of data obtained from the UK Music Charts. The exclusion of music types that would not have been included in the UK Music Charts was conducted manually by the researcher. Any tunes for which the genre was unclear (e.g., the researcher was unfamiliar with the tune) were obtained via Internet search (e.g., YouTube or iTunes recordings) and categorized accordingly. Since the aim of this step was to compile a list of all tunes that could have possibly appeared in the music charts, only tunes that were clearly from non-popular music genres (e.g., classical music) were excluded at this point. Overall, this data compilation process resulted in 3,806 usable responses. In total, 410 of the 3,000 participants (13.67%) did not answer either of the relevant survey questions; all other participants answered at least one of the two questions.

Next, information on each INMI song’s popularity and recency was acquired. These variables were measured in terms of the number of weeks the song had spent in the UK Music Charts (popularity measure), the highest position the song had attained in the charts (popularity measure), and the date the song had exited the charts (recency measure).² Not all 3,806 songs from the first round of data compilation had been in the charts; the resulting number of songs from the original sample that had been listed in the charts was 1,558. From this list, the most frequently mentioned song was named 33 times as INMI (Lady Gaga’s

² This information was acquired via the UK Music Charts database at polyhex.com. These records include songs that were listed in the charts from 1952 to the present date. The exit date variable was converted from a date to the number of days since exiting the charts for use in subsequent analyses. This number of days was calculated from February 22, 2013—the end date of data collection for the project.

“Bad Romance”). The nine most frequently named INMI tunes are listed in Table 1. In this reduced dataset, 1,144 songs were named once as INMI and 414 songs were named more than once (see Figure 1).

Table 1

Songs Most Frequently Named as INMI

| Song Title & Artist | # of Times Named as INMI | Highest Chart Position | Weeks in Charts | Days Since Chart Exit |
|---------------------------------------------------|-----------------------------|---------------------------|--------------------|--------------------------|
| 1) Bad Romance- Lady Gaga | 33 | 1 | 47 | 1322 |
| 2) Can't Get You Out of My Head- Kylie Minogue | 24 | 1 | 25 | 4164 |
| 3) Don't Stop Believing- Journey | 21 | 6 | 59 | 1399 |
| 4) Somebody That I Used to Know- Gotye | 19 | 1 | 46 | 398 |
| 5) Moves Like Jagger- Maroon 5 | 17 | 2 | 52 | 545 |
| 6) California Gurls- Katy Perry | 15 | 1 | 26 | 1083 |
| 7) Bohemian Rhapsody- Queen | 14 | 1 | 17 | 13621 |
| 8) Alejandro- Lady Gaga | 12 | 7 | 10 | 1175 |
| 9) Poker Face- Lady Gaga | 11 | 1 | 66 | 1490 |

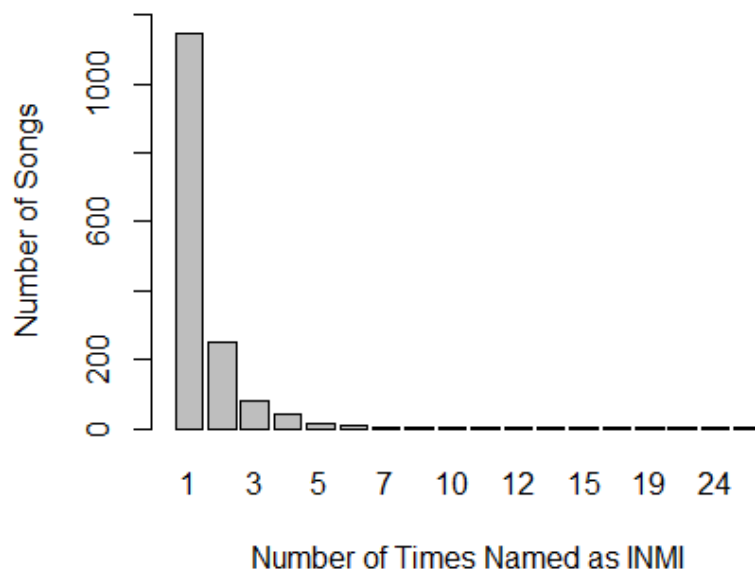


Figure 1. Count data for number of times a song was named as INMI in the “Earwormery” questionnaire (songs with chart data only, $N = 1,558$).

Preliminary Analysis: Predicting INMI Count Based on Popularity and Recency

As a first step, an analysis was conducted to identify the degree to which the popularity and recency of a song could affect the likelihood of the song becoming stuck in the mind as INMI. This primary aim of this analysis was to identify popularity-related variables that contribute to the INMI experience, in order to control for these variables in subsequent analyses of the musical features of INMI tunes.

The data for this analysis require a class of statistical techniques that can model typical distributions of count data. Poisson regression is the most common method for modelling count data that assume a distribution similar to that displayed in Figure 1 (Cameron & Trivedi, 2013; Hilbe, 2011). In the present work, a Poisson regression model was fitted, as well as several Poisson-related models, which were tested as potentially better fits to the data due to both the presence of over-dispersion in the data and the large number of

INMI tunes that were named only one time within the dataset.³ As the Poisson and related models are designed to address data in which the count distribution begins at zero, the present data was transformed by subtracting 1 from each song count (so that the 1-counts for the 1,144 songs named once became zeros, the 2-counts became ones, and so on).

As the two variables that described a song's popularity (highest chart entry and weeks in the charts) were highly correlated ($r(1556) = -.51, p < .001$), these two variables were subjected to a one-component principal component analysis (PCA). The highest chart entry variable loaded negatively onto this component and the weeks in the charts variable loaded positively onto the component. The component scores from this PCA were extracted for use in subsequent regression models as a composite measure of the two popularity variables. The recency variable (days since exiting the charts) was subjected to a square root transformation due a non-normal distribution.

For each of four Poisson and Poisson-related models, a model was fitted using the recency variable and the component scores from the PCA of the popularity variables as predictors of the number of times a song was named as INMI (INMI count). The summaries of these four models are presented in Table 2.

³ One assumption of the Poisson distribution is that the observed mean and the observed variance of the data are equal. However, the observed variance of the present data (2.89) for the song counts is substantially greater than the observed mean (1.55), which indicates over-dispersion. Over-dispersion is a common problem in count data, and can be dealt with by adopting a negative binomial regression model. Additionally, a related family of models—the hurdle and zero-inflated models—were tested as possible models for the data, due to their ability to account for excess zero-counts (Mullahy, 1986; Zeileis, Kleiber, & Jackman, 2007).

Table 2

Parameters and Model Summaries for the four statistical models predicting the number of times a song was named as INMI (INMI count).

| | | Estimate | S.E. | z-value | p-value |
|------------------------------------|----------------|----------|---------|---------|-----------|
| Poisson | Intercept | 0.123 | 0.082 | 1.493 | .135 |
| | Popularity | 0.548 | 0.038 | 14.576 | < .001 ** |
| | Recency (sqrt) | -0.011 | 0.000 | -11.432 | < .001 ** |
| Negative | Intercept | -0.039 | 0.149 | -0.262 | .794 |
| Binomial | Popularity | 0.467 | 0.058 | 5.134 | < .001 ** |
| | Recency (sqrt) | -0.009 | 0.002 | -5.134 | < .001 ** |
| Hurdle: -Count component | Intercept | -6.815 | 51.424 | -0.133 | .895 |
| | Popularity | 0.603 | 0.103 | 5.845 | < .001 ** |
| | Recency (sqrt) | -0.013 | 0.003 | -4.681 | < .001 ** |
| | Log(theta) | -8.415 | 51.442 | -0.164 | .870 |
| -Hurdle component | Intercept | -0.695 | 0.158 | -4.411 | < .001 |
| | Popularity | 0.310 | 0.060 | 5.204 | < .001 ** |
| | Recency (sqrt) | -0.004 | 0.002 | -2.320 | .020 * |
| Zero-Infl.: -Count component | Intercept | -0.039 | 0.140 | -0.280 | .780 |
| | Popularity | 0.467 | 0.056 | 8.395 | < .001 ** |
| | Recency (sqrt) | -0.009 | 0.002 | -5.529 | <.001 ** |
| | Log (theta) | -1.015 | 0.094 | -10.826 | < .001 |
| -Zero- inflated component | Intercept | -9.750 | 105.684 | -0.092 | .926 |
| | Popularity | -0.249 | 29.685 | -0.008 | .993 |
| | Recency (sqrt) | -0.013 | 1.026 | -0.012 | .990 |

Note: * = a significant predictor at the level of $p < .05$, ** = a significant predictor at the level of $p < .001$. The variable named here as "Popularity" represents the component scores from the PCA of the two popularity variables: highest chart entry and weeks in the charts. Hurdle model: Count component models positive counts only, Hurdle component models zeros vs. positive counts. Zero-inflated model: Count component models all of the data, Zero-inflated component models zeros vs. positive counts.

The four models were then compared in terms of the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and log-likelihood (see Table 3). These

three values provide measures of the goodness-of-fit of each model, with smaller values of the AIC and BIC criteria and higher values of the log-likelihood indicating better model fits. Note that the absolute values of these three criteria do not have any meaningful interpretation but depend on the sample size and complexity of the models. Rather, the difference between the criterion values of different models should be interpreted. A difference of at least 3 for the AIC and BIC is often interpreted as meaningful or “positive” (see e.g. Raftery, 1995) and larger differences indicate stronger empirical support for the assumption that one model fits the data better than the other. For models that are constructed as mixtures of individual models (e.g., zero-inflated and hurdle model) there is ambiguity as to how many free-parameters are contained in the model and hence the information criteria are not defined without problems in these cases (and thus are N/A in these cases).

Table 3

Goodness-of-Fit Measures for All Potential Models

| | df | AIC | BIC | Log-Likelihood |
|---------------|----|----------|----------|----------------|
| Poisson | 3 | 3586.935 | 3602.989 | -1790.468 |
| Neg. Binomial | 4 | 2843.822 | 2865.226 | -1417.911 |
| Hurdle | 7 | 2834.526 | NA | -1410.263 |
| Zero-Inflated | 7 | 2849.822 | NA | -1417.911 |

Note: df= degrees of freedom, AIC= Akaike Information Criterion, BIC= Bayesian Information Criterion

The hurdle model appears on several levels to be the most parsimonious solution for modelling the present data. It achieves the lowest AIC value and highest log-likelihood and is

able to account for both the over-dispersion and excess zero-counts present in the data. This model includes both the popularity and recency variables as significant predictors of the number of times a tune was named as INMI. Specifically, songs that had attained higher chart positions and longer runs in the charts and songs that had exited the charts more recently were named as INMI more frequently than less successful and less recent songs. These results suggest a key contribution of features related to the commercial success of a song to the generation of an INMI experience, thus indicating the need to account for the effects of song popularity and recency when investigating the role of features of melodic structure in the occurrence of INMI experiences. More specifically, in subsequent analyses we take song popularity and recency into account by matching INMI tunes to a “control group” of non-INMI tunes in terms of both of these important factors.

Matching INMI and Non-INMI Tunes

A subset of the 1,558 songs with chart data was taken for subsequent analyses that 1) were named as INMI by at least three separate questionnaire participants and 2) had a corresponding high-quality MIDI transcription available from the Geerdes MIDI music database.⁴ Step 1 was included to ensure that the songs included in the present sample had “INMI quality” that transferred across multiple participants; this step reduced the song sample size to 163. Step 2 was included as MIDI transcriptions were required as input for the computational analysis of the melodic features of each song. After this step, a dataset of 129 INMI tunes remained.

A “control group” of non-INMI tunes was compiled for matching to the INMI tunes. In order to ensure that these tunes would provide close matches to all or most INMI tunes, songs by the same or similar performers to the 129 INMI tunes were purposely sought out, as

⁴ <http://www.geerdes.com/>

well as songs from similar time periods, chart positions, and genres to the INMI tunes. The primary constraint imposed on this non-INMI tune dataset was only tunes that had *never* been named as INMI by any of the 3,000 questionnaire participants could be included. This compilation process resulted in a dataset of 438 tunes that had never been named as INMI. The INMI and non-INMI tunes were then subjected to a non-parametric multivariate matching procedure based on a genetic search algorithm (Diamond & Sekhon, 2005; Sekhon & Grieve, 2012), which was implemented in the GenMatch function from the R package “Matching” (Sekhon, 2011). A major benefit of using this automated matching procedure (over more traditional, manual matching methods) was that the two sets of songs could be matched on the basis of five variables at once. The variables used for this matching procedure were: highest chart entry, weeks in the charts, number of days since exiting the charts, artist, and genre. Matching was performed without replacement in order to provide one-to-one pairing of INMI and non-INMI tunes. The caliper argument was set to 1.2 standard deviations, i.e., the values for all of the continuous variables listed above for each non-INMI tune were required to be within 1.2 standard deviations of the values of those same variables for its matched INMI tune. Overall, this matching procedure was able to generate matches for 101 of the INMI tunes, while 28 INMI tunes could not be matched to suitable control tunes within the parameters specified above. Six of the nine top-named INMI songs (66.67%) from Table 1 survived the matching analysis, with no suitable match found for “Don’t Stop Believing,” “Moves Like Jagger,” or “Poker Face.”

Melodic Feature Extraction

MIDI transcriptions for all 202 songs (INMI and non-INMI tunes) were obtained from the Geerdes MIDI music database. The Geerdes database contains high-quality MIDI transcriptions of over 33,000 pieces of music, including the melody line and all

accompaniment/additional instrumentation. The melody line of the section of the song (e.g., chorus, verse, instrumental) reported as INMI by the participants in the “Earwormery” questionnaire was extracted for all 101 INMI tunes. If more than one section was reported by different participants, the section that was reported most frequently was extracted. For cases in which no particular song section was reported, the chorus was extracted, as this is the section of a song that is most commonly reported to be experienced as INMI (Beaman & Williams, 2010; Hyman et al., 2013). The chorus of each of the 101 non-INMI tunes was extracted for comparison to the INMI tune excerpts. The full melody line of each of the 202 songs (including all verses, repetitions, etc.) was also extracted for use in some of the subsequent analyses. During this process, it was noted that one INMI tune (Funky Cold Medina) was comprised primarily of spoken words rather than a melody line. As such, this song and its matched non-INMI tune were excluded from subsequent analysis, leaving a dataset of 100 INMI and 100 non-INMI tunes. All MIDI files were then converted to a textual tabular file format that contains pitch and onset information for each melodic event using the software MELCONV (Frierler, 2005). The melodic data of all 200 songs was then analysed using the FANTASTIC melodic feature extraction software (Müllensiefen, 2009; Müllensiefen & Halpern, 2014).

A total of 82 melodic structural features were computed for each melody using FANTASTIC. These included both first-order and second-order features of the melodies. Most first-order FANTASTIC features have a second-order counterpart. For instance, *p.range* is a first-order feature that calculates the pitch range of a melody. The second-order counterpart of this feature is *dens.p.range*, which compares the pitch range of the melody in question to the pitch ranges of all melodies in the reference corpus and computes the probability density of the pitch range value of a particular melody in comparison to a reference corpus of melodies. The reference corpus used for computing second-order features

in the present study was a collection of 14,063 MIDI transcriptions representative of commercially successful Western pop songs from the Geerdes MIDI music database (Müllensiefen, Wiggins, & Lewis, 2008).

In addition to the distinction between first- and second-order features, several of the features in FANTASTIC are classified as “m-type features”. These are features that aim to capture the usage and repetition of melodic motives in a phrase by taking account of the note order. M-types are calculated within FANTASTIC through the use of a “moving window” that slides over small sections of notes in a melody and records the content of each position of the window, similar to n-gram models that are often used in computational linguistics (e.g., Brown, deSouza, Mercer, Della Pietra, & Lai, 1992). In the present analysis the window was set to vary in size from containing two notes up to six notes at a time (the default setting in FANTASTIC). Examples of m-type features include mean m-type entropy (*mean.entropy* in FANTASTIC), which calculates the average entropy value across all computed m-types for a melody, and mean Yules K (*mean.Yules.K* in FANTASTIC), which is a feature taken from linguistics that measures the rate at which words are repeated in a text—or, in this case, the rate at which musical m-types are repeated in the melody.

In the present analysis, all second-order m-type features were computed using the full melody version of the INMI and non-INMI tunes, in order to take account of the repetition of motives throughout each song as a whole. All other FANTASTIC features were computed using the song excerpts as input. Finally, the tempo of each song excerpt was added to the dataset as an additional predictor of interest, resulting in a total of 83 predictor variables for the subsequent analyses. Tempo information, in beats per minute (bpm), for each song excerpt was obtained from the Geerdes MIDI music database.

Results

The main analysis was conducted using a data classification method known as the random forest (Breiman, 2001), in which the aim was to classify INMI versus non-INMI tunes based on their melodic features. The random forest method has several advantages over other data classification procedures, as it can handle a large number of predictor variables and can cope with non-linear relationships between variables (see Hastie, Tibshirani, Friedman, & Franklin, 2005 for a general description of random forests in classification tasks and data mining, see Strobl, Malley, & Tutz, 2009 for applications of random forests in psychology, and Pawley & Müllensiefen, 2012 for the use of random forests in music psychology). This method also provides a measure of variable importance for each of the predictor variables, which is useful for selecting the most predictive of a large number of variables for further evaluation. Additionally, random forests can model complex higher order interactions between variables that may be difficult to capture using traditional regression methods. The random forest method was implemented in the present research with the “party” package in R (Hothorn, Hornik, & Zeileis, 2006).

A random forest model was first fitted including all 83 predictor variables and INMI versus non-INMI tune as a binary response variable in order to select a subset of variables with the most predictive power.⁵ A variable importance score was obtained for each melodic feature variable, which describes how predictive this variable is in comparison to the other variables. The variable importance scores for the top 12 performing variables are visualized in Figure 2 (see Appendix for a description of these 12 variables).

⁵ Parameters were set using the `cforest_unbiased()` function in R to ensure that the variable selection and variable importance values were unbiased (Strobl, Boulesteix, Zeileis, & Hothorn, 2007). The total number of trees to be grown was set to 1,000, the number of randomly selected variables considered at each split was set to 20, and the minimum number of observations per node necessary for splitting was set to 5.

A “confidence interval” criterion was applied in order to select the top performing variables for further analysis. This criterion specified that only the variables whose (positive) variable importance scores were greater than the absolute value of the lowest negative variable importance score (from the worst performing predictor) would be chosen (Strobl et al., 2009). Only the top three performing variables had a variable importance score that met this criterion (*dens.step.cont.glob.dir*, *tempo*, and *dens.int.cont.grad.mean*). As such, a reduced random forest model was fitted including only these three predictor variables.

After fitting this reduced model with three predictors, a leave-one-out cross validation was performed to provide an unbiased assessment of the model’s classification accuracy. The cross-validated classification rate for classifying INMI versus non-INMI tunes was 62.5%.

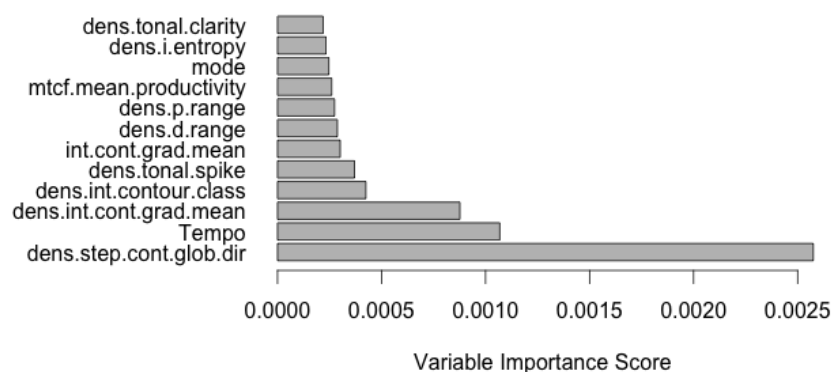


Figure 2. Variable importance scores for the 12 most important predictors in the random forest model (see Appendix for descriptions of each predictor).

Next, the directionality of the relationships between the three variables included in the final random model forest and the dependent variable were investigated, as this information is not directly available from the random forest output. As such, these three variables were entered into a classification tree (see Figure 3) and a binary logistic regression analysis was performed (see Table 4). The classification tree indicated that tunes with a common global

melodic contour in comparison to the corpus of pop songs⁶ were more likely to become INMI (*dens.step.cont.glob.dir* values greater than 0.326); approximately 80% of the tunes fulfilling this criterion were named as INMI tunes by the participants. For tunes with a global melodic contour value less than or equal to 0.326, the feature *dens.int.cont.grad.mean* was then taken into account by the classification tree. This feature measures the commonness of the average gradient of the melodic lines between turning points within the contour.⁷ Tunes that are less similar to the corpus in terms of this features (those with *dens.int.cont.grad.mean* values less than or equal to 0.421) were more likely to become INMI than tunes that have a more common average contour gradient (tunes with *dens.int.cont.grad.mean* greater than 0.421). Within this particular model tempo was not selected as a deciding variable, however this single classification tree should not be interpreted as representing the outcomes of the full random forest, which averages across a large number of classification trees. As such, the results of the binary logistic regression add further clarity to the question of directionality of effects. This regression model indicated that INMI tunes had faster tempi ($M = 124.10$ beats per minute (bpm), $SD = 28.73$) than non-INMI tunes ($M = 115.79$ bpm, $SD = 25.39$) and also indicated the same directionality of effects of the contour variables as the classification tree, although only the *dens.step.cont.glob.dir* variable reached the conventional significance level of $p < .05$. No significant two- or three-way interactions were found between any of the predictor variables in the logistic regression analysis.

⁶ See Figure 4 for examples of common and uncommon global melodic contours.

⁷ See Figure 5 for examples of common and uncommon average gradient between melodic turning points.

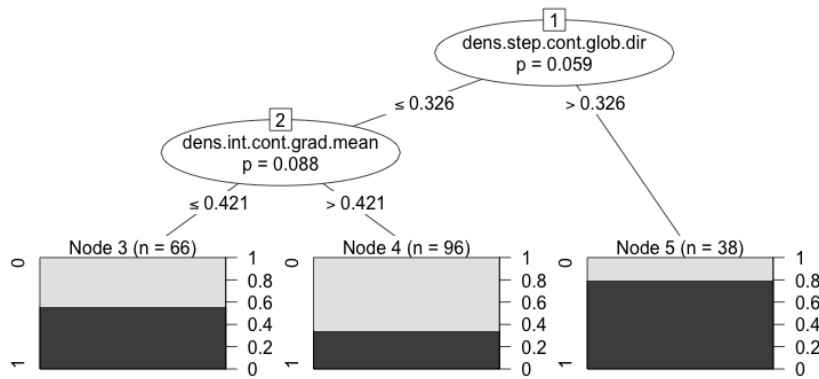


Figure 3. Classification tree resulting from the three predictors used in the random forest.

Table 4

Binary Logistic Regression Analysis Using the Three Random Forest Predictors

| | Estimate | S.E. | z-value | p-value |
|-------------------------|----------|------|---------|---------|
| Intercept | -2.42 | 1.28 | -1.89 | .06 |
| Tempo | 0.01 | 0.01 | 1.66 | .10 |
| dens.int.cont.grad.mean | -1.93 | 1.07 | -1.81 | .07 |
| dens.step.cont.glob.dir | 7.08 | 3.31 | 2.14 | .03* |

Note: * = significant predictor at the level of $p < .05$.

Predicting INMI from Both Melodic Features and Chart Data

As a final step in exploring the factors that increase the likelihood that a tune will be reported as INMI, a combined analysis was conducted that included both the popularity and recency variables from the UK Music Charts *and* the predictions generated by the random forest model of the three most important melodic features. This combined analysis was conducted using the dataset of 100 INMI and 100 matched non-INMI tunes.

As in the previous analysis, the highest UK chart entry and weeks in the charts variables were highly correlated. These variables were thus subjected to a one-component principal component analysis and the component scores were extracted for use as a combined measure of these two variables. The highest chart entry variable loaded negatively onto this component and the weeks in the charts variable loaded positively onto the component. A square root transformation was also applied to the recency variable (number of days since a song exited the charts). Thus, three predictor variables were included in the subsequent analysis: the combined measure of the two popularity variables, the square root-transformed recency variable, and the binary predictions of the final random forest model that made use of three melodic features (*tempo*, *dens.int.cont.grad.mean*, and *dens.step.cont.glob.dir*). The dependent variable of interest was the number of times a tune had been named as INMI. Non-INMI tunes were coded as a value of 0 for this variable. As this dataset of 100 INMI tunes included only INMI tunes that had been named at least 3 times by questionnaire participants, the count variable for each INMI tune was transformed by subtracting 2 from each count, so that there was no gap in the distribution between the zero-counts of the non-INMI tunes and the count data for the INMI tunes.

A hurdle model again provided the best fit to the data, in comparison to related models, based on several goodness-of-fit criteria (see Table 5). The hurdle model resulting from the combined analysis is presented in Table 6. The results indicate that the popularity and recency variables were significant predictors in the count component of the model, which models the positive (non-zero) counts from the dataset, whereas the random forest predictions were a significant predictor in the hurdle component, which models the zero-counts against the larger (non-zero) counts. This result can be interpreted to indicate that the melodic feature data from the random forest significantly predicts whether a tune is named as INMI (whether

its INMI count is zero or non-zero), while the popularity and recency of a song serves to predict *how many times* a tune is named as INMI (for non-zero INMI counts).

Table 5

Goodness-of-Fit Measures for All Potential Models: Combined Analysis

| | df | AIC | BIC | Log-Likelihood |
|---------------|----|--------|--------|----------------|
| Poisson | 4 | 736.18 | 749.37 | -364.09 |
| Neg. Binomial | 5 | 583.70 | 600.19 | -286.85 |
| Hurdle | 9 | 543.22 | NA | -262.61 |
| Zero-Inflated | 9 | 568.51 | NA | -275.26 |

Table 6

Hurdle Model with Extra- and Intra-Musical Predictors of INMI Likelihood

| | | Estimate | S.E. | z- value | p-value |
|-----------|----------------|----------|-------|----------|-----------|
| -Count | Intercept | 0.712 | 0.697 | 1.021 | .307 |
| component | Popularity | 0.629 | 0.162 | 3.885 | .0001** |
| | Recency (sqrt) | -0.012 | 0.005 | -2.554 | .011* |
| | RF Predictions | -0.101 | 0.370 | -0.273 | .785 |
| | Log(theta) | -1.037 | 0.911 | -1.138 | .255 |
| -Hurdle | Intercept | -0.932 | 0.443 | -2.105 | .035 |
| component | Popularity | 0.111 | 0.170 | 0.651 | .515 |
| | Recency (sqrt) | -0.001 | 0.005 | -0.227 | .820 |
| | RF Predictions | 2.302 | 0.336 | 6.842 | < .0001** |

Note: * = a significant predictor at the level of $p < .05$, ** = a significant predictor at the level of $p < .001$. The variable named here as "Popularity" represents the component scores from the PCA of the two popularity variables: highest chart entry and weeks in the charts.

A likelihood ratio test between the hurdle model presented in Table 6 (log-likelihood = -262.61) and an analogous model that includes only the popularity and recency variables (log-likelihood = -290.37) revealed a significant difference between the two models, $X^2(2) = 55.52$, $p < .001$, indicating that the inclusion of the melodic feature-based predictor (predictions from the random forest) provides substantial explanatory power over the model including only song popularity and recency measures. Thus, it appears that both features of a melody and the song's relative success in the charts contribute to the likelihood that a tune is reported as INMI.

Discussion

The results of the present work indicate that features of a song's melodic structure, as well as measures of its popularity and recency, can be useful in predicting whether a song becomes INMI. These findings contribute to the growing literature on the INMI experience and serve to increase our general understanding of why certain songs are spontaneously recalled in the mind over others.

The final random forest model made use of only three melodic features, specifically song tempo and two second-order features that express how common the contour of a song is with respect to the reference corpus. Tunes that were more likely to become INMI were generally faster in tempo than non-INMI tunes, although this predictor was only marginally significant in the logistic regression analysis. Future research could investigate the related question of whether tunes from tempo ranges that are more easily entrained to are more likely to become INMI, particularly considering that a large proportion of INMI episodes reported by participants in previous diary studies occurred during repetitive movements, such as walking or running (Jakubowski et al., 2015).⁸ A second key FANTASTIC feature in the present work was *dens.step.cont.glob.dir*. The findings in relation to this feature indicate that tunes with more common global melodic contour shapes (in terms of the patterns of rising and falling of pitches) are more likely to become INMI than those with less common pitch contours. Some examples of tunes from the present dataset with the least and most common global contour shapes with respect to the reference corpus are provided in Figure 4. In these particular examples, the tunes with more common global contours (section B of Figure 4)

⁸ The mean tempo for INMI and non-INMI tunes from the present work are both close to the average spontaneous motor tempo and preferred perceptual tempo for adults of approximately 120 bpm (McAuley et al., 2006; McKinney & Moelants, 2006). However, future experimental research that attempts to induce INMI episodes using songs from a wider tempo range could shed further light on the relationship between INMI, preferred tempo, and concurrent movement.

assume fairly arch-shaped phrases. This is in line with previous research citing the melodic arch as one of the most common contour shapes in Western music traditions (Huron, 1996). Tunes in the present dataset with less common global contours (section A of Figure 4) appear to take on contours other than arch shapes, such as ascending melodic lines that do not descend again. Finally, the findings related to the feature *dens.int.cont.grad.mean* indicate that tunes with a *less common* average gradient (slope) of the melodic lines between contour turning points are more likely to become INMI. Turning points in a melody occur when the pitch direction reverses (e.g., pitches were ascending and then switch to a descending pattern or vice versa); the gradient of each melodic line is calculated based on how far and how quickly the pitches ascend or descend. Some examples of the extreme values of this particular variable within the present dataset are presented in Figure 5. The tunes with more common average contour gradients within this sample appear to comprise mostly stepwise intervallic motion or repetitions of the same note, whereas the tunes with less common average contour gradients tend to contain many melodic leaps (as in A1 of Figure 5) or unusually large melodic leaps (as in A2 of Figure 5). However, further research is needed to examine whether such contour patterns hold across other musical genres that are experienced as INMI (e.g. classical music).

In sum, tunes that become INMI tend to be faster in tempo than non-INMI tunes. If the melodic contour shape of a melody is highly congruent with established norms, then it is more likely for the tune to become INMI. If the melodic contour does not conform with norms, then it should have a highly unusual pattern of contour rises and falls to become an INMI tune.

A) Uncommon Global Contours (lowest values of *dens.step.cont.glob.dir*):



A1) Owner of a Lonely Heart (Yes) (INMI tune)



A2) Rock 'N' Me (Steve Miller Band) (non-INMI tune)

B) Common Global Contours (highest values of *dens.step.cont.glob.dir*):



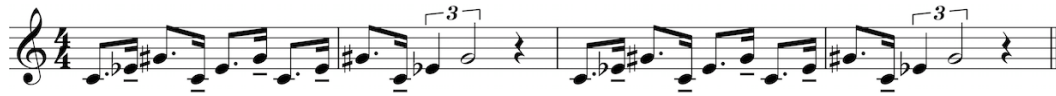
B1) Smoke on the Water (Deep Purple) (INMI tune)



B2) Plug in Baby (Muse) (INMI tune)

Figure 4. Examples from the present dataset with the lowest (A) and highest (B) values of the variable *dens.step.cont.glob.dir*.

A) Uncommon Avg. Contour Gradient (lowest values of *dens.int.cont.grad.mean*):



A1) In the Mood (Glenn Miller) (INMI tune)



A2) My Sharona (Knack) (INMI tune)

B) Common Avg. Contour Gradient (highest values of *dens.int.cont.grad.mean*):



B1) Lucky Man (Verve) (non-INMI tune)



B2) Intergalactic (Beastie Boys) (INMI tune)

Figure 5. Examples from the present dataset with the lowest (A) and highest (B) values of the variable *dens.int.cont.grad.mean*.

The melodic features that were most predictive of the likelihood of a tune to become INMI bear some relation to previous literature. The “Hooked” project on musical catchiness has revealed audio features related to melodic “conventionality” as positive predictors of the long-term salience of a melody (Van Balen et al., 2015), which bears some conceptual similarity to the finding in the present work that INMI tunes generally comprised more common global melodic contours than non-INMI tunes. Müllensiefen and Halpern (2014) reported a variety of features that predicted explicit and implicit memory for previously unfamiliar songs. Interestingly, there is little overlap between the features revealed by their research and the present study, with the possible exception of the fact that they found simple contours to be predictive of implicit melodic memory. However, as Müllensiefen and Halpern’s study investigated recall of novel tunes after only a single exposure it is not surprising that the melodic features implicated in their work are rather different to the present findings on features that enhance the spontaneous recall and repetition of (often highly) familiar music within musical imagery. Additionally, the finding in the present work that INMI tunes tended to be faster in tempo than non-INMI tunes does not appear to have a previous precedent in existing literature. It is of interest to explore this finding further in terms of potential relationships of this tempo variable to sensorimotor and entrainment processes.

The present work did not replicate the findings of Finkel et al. (2010) and Williamson and Müllensiefen (2012), who reported that, on average, INMI tunes made use of longer note durations and smaller pitch intervals than non-INMI tunes. The present findings may be at least somewhat related to Williamson and Müllensiefen’s interpretation that INMI tunes may be easier to sing, as, for instance, common global contours may be easier to sing than less common contours. However, the work of Finkel et al. (2010) and Williamson and

Müllensiefen (2012) did not make use of second-order features that combine information from the individual melodies analyzed with information from a large reference corpus of melodies. In the present study, the two melodic contour features selected by the random forest were both second-order features. The fact that none of the first-order features selected in previous studies with much smaller sample sizes were selected in the present study, which included both first- and second-order features, seems to point to the importance of taking account of corpus information regarding the distribution of features values, i.e. their commonness/rarity. In addition, the present results are likely more reliable than the initial results reported in Finkel et al. (2010) and Williamson and Müllensiefen (2012), as the present work used a larger sample of tunes and a more controlled procedure for matching INMI and non-INMI tunes. The use of the random forest method also confers several advantages over the logistic regression method used in the previous studies. For instance, the random forest can easily model interactions between multiple variables and allows for different “earworm formulas” to be modelled within the different trees of the forest that do not have to have much in common with one another. Given the wide diversity of tunes that are included within the genre of modern “pop music”, this multiple formula hypothesis seems to be more plausible than a single formula common to all INMI tunes.

The present work has also revealed that features related to a song’s popularity and recency—in particular, the song’s highest UK Chart entry, the number of weeks the song spent in the charts, and the number of days since the song has exited the charts—can play a significant role in predicting the *number of times* a song is named as INMI by participants. These findings share some similarities with previous work on extra-musical features that influence INMI occurrence. For instance, Byron and Fowles (2013) reported that previously unfamiliar songs were more likely to become INMI if participants were exposed to them six rather than two times, thereby suggesting a role of familiarity in the INMI experience.

Additionally, recency of exposure can play a key role, such that songs that have been heard aloud more recently are more likely to become INMI than songs that were heard less recently (Hyman et al., 2013; Liikkanen, 2012b). The UK chart variables used in the present work are conceptually related to familiarity and recency measures, as songs that achieve higher positions and longer runs in the charts may be played more often on the radio and online platforms such as YouTube and Spotify, thereby increasing their familiarity in the listener. Songs that have more recently been in the charts may also have a greater chance of having been recently heard by participants than songs that were in the charts many years ago. Future research might also consider the recency variable (days since exiting the charts) in relation to the age of the participant who reported a tune as INMI. This would allow for the exploration of questions such as whether participants more frequently report INMI for songs that were released during certain period of their lives, such as the “reminiscence bump” (a period in late adolescence/early adulthood from which autobiographical memories tend to be disproportionately recalled; Rubin, Wetzler, & Nebes, 1986).

There are several potentially promising avenues for future investigation and expansion of the present research that should be highlighted here. First, the present work comprises only symbolic data analysis and does not include measures of audio features derived from the actual song recordings, such as loudness, timbral content, rhythmic clarity, etc. Additionally, there may be other key compositional features not represented within the single-line melodic analysis implemented in FANTASTIC, such as the harmonic content or chord structure of the music, articulation, and expressive timing, which could contribute to the INMI nature of a tune. An expanded version of the present study that includes analysis of audio features as well as other structural features of the INMI tunes would be highly beneficial in terms of identifying additional features that can increase the classification accuracy of the models. It should also be noted that the reference pop music corpus that was

available for the analysis of second-order features in FANTASTIC (Müllensiefen et al., 2008) comprises only songs composed before 2007; as such, future efforts should be made to update this reference corpus to include more recently-composed songs to capture any general stylistic changes that may have occurred in the pop music genre since 2007. Finally, an analysis of the lyrical content of INMI tunes could be beneficial, in terms of investigating whether linguistic features, such as rhyme or alliteration, play a role in increasing the likelihood of a song toward becoming INMI. Future research should also compare the results of the current study to data from music styles not included in the present sample in order to identify whether the melodic features of INMI revealed in the present work may be genre invariant.

To summarize, the outcomes of the present research indicate that certain musical features of a tune, as well as measures of its chart success, can be used to predict the likelihood that the song will become INMI. The results of this work may be of interest to researchers of musical and involuntary memory, as well as to music composers and advertisers interested in writing music that will continue to be spontaneously replayed in one's head long after the initial music exposure period. It is possible that the melodic features revealed in this work as predictors of INMI might serve more general functions in terms of increasing the ease with which a tune can be retrieved from memory, although further research needs to be conducted to test this possibility. As the present findings indicate a role of both melodic features and popularity/recency of a song in the genesis of an INMI experience, it would be highly beneficial in future work to begin to construct models that take account of not only acoustic, melodic, harmonic, and lyrical features of melodies, but also participant-level factors such as listening histories, personal associations with the music, and endogenous states (e.g., mood), to provide a more comprehensive account of the factors that contribute to the onset of an INMI episode.

References

- Bailes, F. (2010). Dynamic melody recognition: Distinctiveness and the role of musical expertise. *Memory & Cognition*, *38*(5), 641–50. doi:10.3758/MC.38.5.641
- Bailes, F. (2015). Music in mind? An experience sampling study of what and when, towards an understanding of why. *Psychomusicology: Music, Mind and Brain*, *25*(1), 58–68.
- Beaman, C.P., & Williams, T.I. (2010). Earworms (stuck song syndrome): Towards a natural history of intrusive thoughts. *British Journal of Psychology*, *101*, 637–653. doi:10.1348/000712609X479636
- Beaman, C.P., & Williams, T.I. (2013). Individual differences in mental control predict involuntary musical imagery. *Musicae Scientiae*, *17*, 398–409. doi:10.1177/1029864913492530
- Beaty, R.E., Burgin, C.J., Nusbaum, E.C., Kwapil, T.R., Hodges, D.A., & Silvia, P.J. (2013). Music to the inner ears: Exploring individual differences in musical imagery. *Consciousness and Cognition*, *22*(4), 1163–1173. doi:10.1016/j.concog.2013.07.006
- Bertrand, F., Magnanensi, J., Meyer, N., & Maumy-Bertrand, M. (2014). plsRglm: Algorithmic insights and applications. Technical Documentation. Retrieved from <https://cran.r-project.org/web/packages/plsRglm/vignettes/plsRglm.pdf>
- Bradford, C. (2005). *Heart & soul: Revealing the craft of songwriting*. London: Sanctuary Publishing Ltd.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324
- Brown, S. (2006). The perpetual music track: The phenomenon of constant musical imagery. *Journal of Consciousness Studies*, *13*(6), 43–62.
- Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., & Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, *18*(4), 467-479.
- Burgoyne, J.A., Bountouridis, D., Van Balen, J., & Honing, H. (2013). Hooked: A game for discovering what makes music catchy. In *Proceedings of the International Society for Music Information Retrieval (ISMIR)*, 245-250.
- Burns, G. (1987). A typology of “hooks” in popular records. *Popular Music*, *6*(1), 1–20. doi:10.1017/S0261143000006577
- Byron, T.P., & Fowles, L.C. (2013). Repetition and recency increases involuntary musical imagery of previously unfamiliar songs. *Psychology of Music*, *43*(3), 375-89. doi:10.1177/0305735613511506
- Cameron, A.C. & Trivedi, P.K. (2013), *Regression analysis of count data* (2nd ed.).

Cambridge, UK: Cambridge University Press.

- Dhanaraj, R., & Logan, B. (2005). Automatic prediction of hit songs. *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, pp. 488–491.
- Diamond, A., & Sekhon, J.S. (2005). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. Technical report, Department of Political Science, UC Berkeley. URL: <http://sekhon.berkeley.edu/papers/GenMatch.pdf>
- Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music*, 34(3), 337–371.
- Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception*, 18(3), 275–296.
- Finkel, S., Jilka, S.R., Williamson, V.J., Stewart, L., & Müllensiefen, D. (2010). Involuntary musical imagery: Investigating musical features that predict earworms. Paper presented at the *Third International Conference of Students of Systematic Musicology (SysMus10)*, University of Cambridge, UK.
- Floridou, G.A., & Müllensiefen, D. (2015). Environmental and mental conditions predicting the experience of involuntary musical imagery: An experience sampling method study. *Consciousness and Cognition*, 33, 472–86. doi:10.1016/j.concog.2015.02.012
- Floridou, G., Williamson, V.J., & Müllensiefen, D. (2012). Contracting earworms: The roles of personality and musicality. In Cambouropoulos, E., Tsougras, C., Mavromatis, K., Pasteriadis, K., (Eds), *Proceedings of ICMPC-ESCOM 12* (pp. 302–310).
- Friedman, J., Hastie, T., Simon, N., & Tibshirani, R. (2015). Package 'glmnet.' Technical Documentation. Retrieved from <https://cran.r-project.org/web/packages/glmnet/glmnet.pdf>
- Frieler, K. (2005). Melody - csv file format (mcsv). Technical Documentation (unpublished).
- Frieler, K., Jakubowski, K. & Müllensiefen, D. (in press). Is it the song and not the singer? Hit song prediction using structural features of melodies. *Yearbook of Music Psychology (Jahrbuch Musikpsychologie)*.
- Gower, J.C. (1971). A General coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857–871. doi:10.2307/2528823
- Halpern, A.R., & Bartlett, J.C. (2011). The persistence of musical memories: A descriptive study of earworms, *Music Perception*, 28(4), 425–432.

- Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2), 83-85.
- Hemming, J. (2009). Zur Phänomenologie des 'Ohrwurms' in W. Auhagen, C. Bullerjahn & H. Höge (Eds.) *Musikpsychologie - Musikalisches Gedächtnis und musikalisches Lernen*. Jahrbuch 20, pp. 184-207. Göttingen: Hogrefe.
- Hilbe, J.M. (2011). *Negative binomial regression* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.
- Hothorn, T. Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651-674.
- Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology*, 10, 3-23.
- Hyman, I.E., Burland, N.K., Duskin, H.M., Cook, M.C., Roy, C.M., McGrath, J.C., & Roundhill, R.F. (2013). Going Gaga: Investigating, creating, and manipulating the song stuck in my head. *Applied Cognitive Psychology*, 27, 204–215. doi:10.1002/acp.2897
- Jakubowski, K., Farrugia, N., Halpern, A.R., Sankarpani, S.K., & Stewart, L. (2015). The speed of our mental soundtracks: Tracking the tempo of involuntary musical imagery in everyday life. *Memory & Cognition*. doi: 10.3758/s13421-015-0531-5
- Kopiez, R. & Müllensiefen, D. (2011). Auf der Suche nach den „Popularitätsfaktoren“ in den Song-Melodien des Beatles-Albums Revolver: eine computergestützte Feature-Analyse. In S. Meine & N. Noeske (Eds.), *Musik und Popularität. Beiträge zu einer Kulturgeschichte zwischen 1500 und heute* (pp. 207–225). Münster: Waxmann Verlag.
- Leikin, M.A. (2008). *How to write a hit song*. Milwaukee, WI: Hal Leonard Books.
- Liikkanen, L.A. (2012a). Musical activities predispose to involuntary musical imagery. *Psychology of Music*, 40, 236–256. doi:10.1177/0305735611406578
- Liikkanen, L.A. (2012b). Inducing involuntary musical imagery: An experimental study. *Musicae Scientiae*, 16(2), 217–234. doi:10.1177/1029864912440770
- McAuley, J.D., Jones, M.R., Holub, S., Johnston, H.M., & Miller, N.S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, 135(3), 348–367. doi:10.1037/0096-3445.135.3.348

- McKinney, M. F., & Moelants, D. (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels?. *Music Perception*, *24*(2), 155-166.
- Müllensiefen, D. (2009). FANTASTIC: Feature ANalysis Technology Accessing STatistics (In a Corpus). Technical Report. URL: http://www.doc.gold.ac.uk/isms/m4s/FANTASTIC_docs.pdf.
- Müllensiefen, D., & Halpern, A.R. (2014). The role of features and context in recognition of novel melodies. *Music Perception*, *31*(5), 418–435. doi:10.1525/mp.2014.31.5.418
- Müllensiefen, D., Jones, R., Jilka, S., Stewart, L., & Williamson, V.J. (2014). Individual differences predict patterns in spontaneous involuntary musical imagery, *Music Perception*, *31*(4), 323–338.
- Müllensiefen, D., Wiggins, G., & Lewis, D. (2008). High-level feature descriptors and corpus-based musicology: Techniques for modelling music cognition. In A. Schneider (Ed.), *Systematic and Comparative Musicology: Concepts, Methods, Findings* (pp. 133-155). Frankfurt: Peter Lang.
- Ni, Y., Santos-Rodríguez, R., Mcvicar, M., & De Bie, T. (2011). Hit song science once again a science? *4th International Workshop on Machine Learning and Music*.
- Pachet, F., & Roy, P. (2008). Hit song science is not yet a science. In: *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 355–360.
- Pawley, A., & Müllensiefen, D. (2012). The science of singing along: A quantitative field study on sing-along behavior in the north of England. *Music Perception*, *30*(2), 129-146.
- Perricone, J. (2000). *Melody in songwriting*. Boston, MA: Berklee Press.
- Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163.
- Rubin, D. C., Wetzler, S. E., & Nebes, R. D. (1986). Autobiographical memory across the adult lifespan. In D.C. Rubin (Ed.), *Autobiographical memory* (pp. 202-221). New York: Cambridge University Press.
- Schulkind, M.D., Posner, R.J., & Rubin, D.C. (2003). Musical features that facilitate melody identification: How do you know it's "your" song when they finally play it? *Music Perception*, *21*(2), 217–249. doi:10.1525/mp.2003.21.2.217
- Sekhon, J.S. (2011). Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software*, *42*(7), 1–52.
- Sekhon, J. S., & Grieve, R. D. (2012). A matching method for improving covariate balance in cost-effectiveness analyses. *Health Economics*, *21*(6), 695-714.

- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*(4), 323–348.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288.
- Van Balen, J., Burgoyne, J.A., Bountouridis, D., Müllensiefen, D., & Veltkamp, R. (2015). Corpus analysis tools for computational hook discovery. *Proceedings of the International Symposium on Music Information Retrieval (ISMIR 2015)*.
- Williamson, V.J. & Müllensiefen, D. (2012). Earworms from three angles. In: Cambouropoulos, E., Tsougras, C., Mavromatis, K., Pasiadis, K., editors. *Proceedings of ICMPC-ESCOM 12* (Thessaloniki, Greece), 1124–1133.
- Williamson, V.J., Jilka, S., Fry, J., Finkel, S., Müllensiefen, D., & Stewart, L. (2012). How do earworms start? Classifying the everyday circumstances of Involuntary Musical Imagery. *Psychology of Music, 40*(3), 259-84. doi:10.1177/0305735611418553
- Williamson, V.J., & Jilka, S.R. (2013). Experiencing earworms: An interview study of Involuntary Musical Imagery. *Psychology of Music, 42*, 653–670. doi:10.1177/0305735613483848
- Williamson, V.J., Liikkanen, L.A., Jakubowski, K., & Stewart, L. (2014). Sticky tunes: How do people react to involuntary musical imagery? *PLoS ONE, 9*(1).doi: 10.1371/journal.pone.0086170
- Wold, S., Ruhe, A., Wold, H., & Dunn, III, W.J. (1984). The collinearity problem in linear regression: The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing, 5*(3), 735–743. doi:10.1137/0905052
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301–320. doi:10.1111/j.1467-9868.2005.00503.

Appendix

Descriptions of each of the 12 FANTASTIC features selected in the initial random forest model (see Figure 2). For formal definitions and more detailed explanations of all features see Müllensiefen (2009).

| FANTASTIC feature | Description |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| dens.d.range | 2nd-order measure of the range of note durations* |
| dens.i.entropy | 2nd-order measure of entropy within the set of pitch intervals* |
| dens.int.cont.grad.mean | 2nd-order measure of the mean of the absolute gradient of the melodic contour arising from interpolation lines between contour turning points* |
| dens.int.contour.class | 2nd-order measure of the overall direction of the interpolation contour; can assume 5 values (strong down, down, flat, up, strong up)* |
| dens.p.range | 2nd-order measure of pitch range* |
| dens.step.cont.glob.dir | 2nd-order measure of the overall direction of the step contour* |
| dens.tonal.clarity | 2nd-order measure of tonal clarity; correlations are computed between the melody and every possible major/minor key. This feature then computes the ratio of the highest key correlation to the second highest key correlation* |
| dens.tonal.spike | 2nd-order measure of tonal clarity, similar to dens.tonal.clarity; computes the ratio of the highest key correlation to the sum of all other key correlations* |
| int.cont.grad.mean | mean of the absolute gradient of the interpolation contour; informs about the degree of inclination at which the interpolation contour is rising or falling on average |
| mode | major or minor tonality |
| mtcf.mean.productivity | 2nd-order measure of repetitiveness of m-types (segments of music 2-6 notes in length)* |
| tempo | tempo in beats per minute (bpm) |

*Note: All 2nd-order features compare the values of the feature computed for a single melody to the values of that same feature computed across the reference corpus (14,063 pop songs).