7

HOOKED

The aim of the next two chapters is to analyze the phenomenon of *hooks*. This chapter presents Hooked, the music game we developed to collect data on the music memory of a large number of frequent popular music listeners in the Netherlands and the United Kingdom. The analysis of the data will be presented in chapter 8.

The first two sections introduce the question that will be addressed, and the necessary backgrounds offered by musicology and music cognition. Sections 7.3 and 7.4 describe the design and implementation of the Internet experiments we set up to collect a dataset of hooks.[1]

7.1 CATCHINESS, EARWORMS AND HOOKS

Music cognition, as introduced in chapter 1, is the science of music listening. It seeks to explain the wide range of processes that are involved when a person listens to music, including memory, attention, expectation and emotion [152].

In the scientific study of popular music, this listener-centered perspective on music and the experiences around it is crucial. Popular music has often been characterized by the length of songs: three to four minutes in most cases. Listening to popular music requires a different kind of attention than, say, symphonic music. Memory is very important too: a popular or particularly engaging pop song is often said to be 'catchy' or 'an earworm', suggesting it has somehow been

1 The names *Hooked!* and *Hooked on Music* will be used to distinguish the implementations from the experiment in its most general form.

etched into memory: popular music is, on a fundamental level, an interaction with music memory.[2]

The notion of catchiness is a fuzzy one and difficult to define, but it is generally understood as 'easily recalled to memory', or 'memorable'. Many listeners confidently discriminate between music they consider catchy, and non-catchy music, even upon first or second listen. In the musicology and music cognition literature, however, the notion of catchiness is rarely analyzed, despite its vital role in music memory and popular music.

There is some literature on the related notions of *hooks* and *earworms*. *Earworms* are songs or parts of songs that get stuck in one's head. Over the last decade, earworms have become a serious subject of research in music psychology and neuroscience, where they are often referred to as involuntary musical imagery, and so have a range of related phenomena, such as musical hallucinations and other, non-musical imagery. Little is known about the kind of songs that easily get stuck in one's head, but we are learning fast about the context in which it happens, and the individual differences that correlate with it [84, 133].

A *hook*, in the songwriting literature and in musicology, is the part of a song that make it catchy. It is the part that grabs the attention; it grabs or 'hooks' the listener, trapping the song into the listeners' memory as the result of its memorability [28]. As with earworms, very little is known about what contributes to the 'memorability' or long term memory salience of popular music, about what makes a hook a hook [187]. In the next section, we define the notion of catchiness and hooks, and review what is known about the musical variables that are associated with the phenomenon.

2 Of course, popular music, on an equally fundamental level, must also be understood in terms of processes of embodied cognition, identity and social dynamics, semiotics... We consider the music's interaction with memory an important facet of the popular music experience, but, recalling Huron's notion on explanatory 'closure', we do not choose the angle of music and memory with the intent of replacing or explaining away other perspectives on popular music—see chapter 1, and [79].

7.2 HOOKS IN MUSICOLOGY AND MUSIC COGNITION

From a cognitive point of view, we can define a hook to be the most salient, easiest-to-recall fragment of a piece of music. Likewise, we can define catchiness as long-term musical salience, the degree to which a musical fragment remains memorable after a period of time. By this definition, every piece of music will have a hook—the catchiest part of the piece—even if some pieces of music 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.

This definition—the hook as the location of maximum catchiness—makes the notion of hooks useful for applications, too. A retrieval system that knows the location of the most memorable or recurring part of each song has a strong advantage in human-centered context of information retrieval, e.g. when the user of a music search system expects the system to return something they know [69]. The system could also benefit computationally. In a similarity search, it might ignore redundant or less salient song sections, thus limiting the space over which it needs to search. In other words, understanding hooks can be useful not only to music cognition, but also to music information retrieval.

We review the available literature on hooks and catchiness in the remainder of this section, beginning with musicological accounts in section 7.2.1, followed by an overview of relevant music cognition perspectives in 7.2.2.

7.2.1 *Hooks in Musicology*

Burns' Typology of Hooks

The most important and the only article-length musicological study of hooks is a 1987 typology of popular music hooks by Gary Burns. [28]. The paper consists of an exhaustive overview of hook categories, extensively illustrated with examples.

Textual elements	Non-textual elements
musical elements:	performance elements:
– rhythm	– instrumentation
– melody	– tempo
– harmony	– dynamics
lyrics	– improvisation
	production elements:
	– editing
	– mix
	– channel balance
	– signal distortions
	– effects

Table 4.: Textual and non-textual elements of a music recording that can make up a hook. Adapted from [28].

Adopting a definition by Monaco and Riordan, Burns starts off by defining a hooks as 'a musical or lyrical phrase that stands out and is easily remembered' [131]. He then proposes that popular music hooks lie in the interaction between *repetition*, *variation* (or *modulation*) and *change* in the elements that make up popular music.

Repetition and change are contrasted to conclude that a discourse entirely based on repetition or continuous change cannot be successful while variation or modulation, terms used interchangeably, are put forward as the critical mechanism by which catchy music is composed.

Burns divides the musical elements that constitute a recording into *textual* elements (melody, harmony, rhythm and lyrics) and *non-textual* elements. The textual elements can be seen as that which can be written, as opposed to non-textual elements which are a product of the performance or production. The set of textual and non-textual elements that could make up a hook is then narrowed down to the list shown in Table 4. The remainder of the paper provides examples.

Not all of the examples are cases of modulation and repetition. Other terms also recur frequently, primarily *distinctiveness*, *surprise* and *intertextuality*. The last term is used to denote hooks that are the result of confusing or associating a piece with a specific other work.

Other work

Other work that theorizes the concept of hooks includes a more recent analysis of hooks by Kronengold [99]. Kronengold suggests that the characteristics of hooks vary across genres, and explores the idea that assortments of hook characteristics might constitute a useful definition of genre.

The analysis further centers on the the possible role of 'musical accidents' (mistakes in performance or production) as hooks. To perceive something as an accident requires a high-level understanding of the music, including the context and the intentions of the performer. Such a strong dependency on context and interpretation places this type of hook alongside the narratives in lyrics and personal associations of the listener, aspects of musical cognition that fall outside the scope of this project.

7.2.2 *Hooks and Music Cognition*

Clearly, Burns' 1987 review is written from a popular music research perspective, and only a few times do notions related to cognition show up in the discussion. What are the implications of Burns' systematization when the discussion is read in light of theories in the music cognition domain? The important keywords in this analysis that are shared with the vocabulary of music cognition are *repetition*, *variation-modulation* and *distinctiveness*. We review the prevailing perspective on these topics.

Repetition and Variation

Elizabeth Margulis has written extensively about the role of *repetition* in Western and non-Western music—identifying repetition as a distinctive property of earworms, for example, and studying the effects on human repetition detection of multiple exposures [117, 118]. Results on the latter suggested that not all repetition is the same: within-phrase repetition is more noticeable than separated, between-phrase repetition . Repetition of units that form a complete segment at some level also seems to be more noticeable than repetition of units that do not.

Results from another study by Margulis suggested that listeners aesthetically prefer slight variations over verbatim repetitions, in two experiments on contemporary Western Art music and eighteenth century Rondos. This supports Burns' initial argument about the appeal of variation as compared to exact repetition [119].

Studying repetition in lyrics, Nunes et al. showed that songs with repeated lyrics have a higher chance of debuting in the Top 40 charts, and an higher chance of reaching the number one position [140]. The correlation is attributed to an increase of processing fluency of the song's lyrics.

Effects of repetition on musical preference also seem consistent with the *mere exposure effect*: listeners tend to show an increased liking of a piece of music as they become more familiar with it. Huron ascribes this effect to our brains rewarding correct predictions: "listeners prefer familiar stimuli not because they are familiar, but because they are predictable" [82]. In other words: we prefer listening to music with repetitions because the repetitions make it predictable. A number of such associations between familiarity and reward have been confirmed in brain imaging studies, e.g., by Pereira et al. [157]. They show that the brain's 'reward circuitry' is more active for song excerpts that were familiar to the participant prior to the experiment. However, we should be careful to generalize from these experiment, as the kind of familiarity studied tend to vary; in [157], it is the result of repeated exposure on the time scales of weeks, months or years, so it is far removed from the familiarity induced by repetitions within a piece of music. The study also shows an increase in reward, not preference.[3]

3 In the part of the experiment based on familiarity ratings, preference ratings was even controlled for, as any relation between preference and familiarity, in a correlational study, is seen as confounding—we tend to be more familiar with the music we like more.

Chart position and preference ratings are not necessarily good measures of catchiness. They are strongly affected by exposure and the many economic, social, demographic and other factors that in turn affect exposure. In other words, the above arguments are worth some consideration, but while the evidence for an effect of familiarity on preference is strong, the evidence for repetition as a driver of preference is incomplete.

Distinctiveness

A study that does aim to measure memorability, is Müllensiefen and Halpern's experiment on the recognition of popular music melodies [134]. Presenting participants with a set of unknown popular song melodies, some once and some more than once, they ask two questions: have you heard this melody before (within the experiment)? And, how would you rate this melody in terms of pleasantness? The two kinds of ratings are used as an explicit and implicit measure of recall, respectively. Evidence from experiments on the mere exposure effect is cited to support the argument that pleasantness ratings can be a more reliable proxy of implicit familiarity than familiarity as measured by 'explicit' ratings. They then investigate the features that predict recognition of the melodies using a 'discovery-driven' statistical analysis. Given a set of features computed from the (symbolic) melodies, a regression model is used to find the latent factors in a set of symbolic features that best predict the two memory scores.

Results show that explicit memory scores ('I have heard this before') are higher for less typical and less complex melodies, as measured by the mean document frequency and productivity of n-grams, respectively (variables from the FANTASTIC toolbox—see [132] for definitions). Implicit memory shares one of these two factors: scores are higher for melodies composed of less typical motives. In terms of complexity, however, the trend is inverted: more complex memories get higher ratings. These findings—for either rating, less typical melodies get higher scores—suggest a link with *distinctiveness*.

In a more general perspective on distinctiveness, Huron presents in [79,80] an overview of the literature on schema selection and expectation in music. Huron highlights three types of expectation:

Schematic expectations
are expectations that arise from the cumulative exposure to music throughout all of one's life. Several schemas may exist in parallel. Upon listening, schemas are selected based on the type of music (eg. its genre) and the context. Schematic expectations have a quick effect.

Veridical expectations
are the expectations that are associated with knowing a particular work. Veridical expectations are slowly triggered as they require a confident recall of the piece.

Adaptive expectations
are dynamically updated and accumulate during listening of the piece itself. They are especially prominent when listening to a work for the first time. First proposed by Meyer [128].

In the context of Müllensiefen's findings in [134], schematic expectations would reflect 'typicality', what is 'common' in popular music. A violation of these expectations occurs whenever atypical, distinctive patterns are encountered. In other words, Müllensiefen provides some evidence for a relationship between distinctiveness and memory salience, perhaps including long term memory salience and hooks. In light of Huron's technical perspective on expectation, this can be read as evidence for an effect of violations of schematic expectation.

These results are valuable clues in the understanding of memorability and catchiness, however, the scope and context of these findings is important, too: they are based on data from a a small group of 34 participants listening to short, synthesized, monophonic melodies. And most importantly, the test only probes recent memory—somewhere on the order of the length of the experiment. The particularities of this kind of musical memory may be quite different from the long-term memory salience of hooks.

7.2.3 *Summary: Hook Types*

The above imperfect, but valuable findings in music cognition point to, on one hand, the importance of repetition, reinforcing adaptive expectations and contributing to greater processing fluency, and on the other hand, distinctiveness (i.e., violations of schematic expectations).

Emphasizing modulation/variation and distinctiveness, Burns' typology allows for both kinds of hooks. As an example of the first, 'vamp' hooks get their catchiness from the continued repetition of a (often very stereotypical) pattern, conforming to schematic expectations, as in e.g., Louie Louie by The Kingsmen (1963), a song built on a short repeated I-IV-V-IV progression. Similarly, many of Burns' examples of instrumentation and production-related hooks are essentially cases of atypical or distinctive sound or timbre, e.g., the use of a sitar in The Beatles' Norwegian Wood (1965).

Alongside these two types, Burns' typology also allows for the combination of both effects: pattern that are repeated, and thereby very representative for the song, but not for the larger body of music the song belongs to, a notion of hooks that may be referred to as an 'identifying motive'. Examples include the repeated I-bII harmonic progression in Jefferson Airplane's White Rabbit (1967).

One recurring and potentially interesting keyword out of Burns' analysis that hasn't been given much attention in the music cognition literature is *surprise*. Surprise is probably most easily understood in terms of Hurons ideas as an effect of adaptive expectations. Adaptive expectations represent the patterns that are established as a song comes along. In this perspective, adaptive expectations may be reinforced by repetition and violated at moments of surprise. A hook that is cited as an example of both distinctive (unusual in absolute terms) and surprising (in relation to the rest of the melody) is Minnie Ripperton's high-pitched vocal effort in Lovin' You (1975).

These many possible types of hooks—repetition, distinctiveness, identifying motive, surprise—directly and indirectly hypothesized in the literature, will be considered in the next chapter. They will not be regarded as strict hypotheses, to be tested individually, but they are the motivation for an analysis in which approximate expectations are summarized in 'corpus-based features' (section 8.1).

7.3 EXPERIMENT DESIGN

In the work described in this section, we set out to collect a dataset of hooks: properly sourced annotations of the catchiest parts of a set of songs, where catchiness is measured in a way that respects its above definition as long-term musical salience.

7.3.1 *Measuring Recognisability*

Collecting a dataset of hooks presents a number of challenges. First, an appropriate measure of a hook's catchiness must be found. Having defined the hook as the most memorable, easiest-to-recall fragment, we can look at the psychology literature for measures of recall.

In memory retrieval experiments in the domain of psychophysics, *two-alternative forced choice tasks* are a common experiment paradigm. In this set-up, a participant is asked to answer a binary-choice question, and the response is timed. When applied in memory research, a range of memory retrieval models is available that allow for the choice and response time of items and participants to be combined into a single measure of ease-of-recall, e.g. Ratcliff's drift-diffusion model—more on this in chapter 8 [164].

Following this paradigm, we choose as an empirical measure of catchiness the *stimulus drift rate*, i.e., the ease-of-recall as measured using the drift diffusion model of memory retrieval based on response times. Admittedly, this does not cover all of the possible associations that the informal and overloaded term 'catchiness' might have (e.g., any association with the notion of earworms), but it gives us a rigorous and practical means of measuring the phenomenon.

7.3.2 *Games and Music Research*

A second challenge is that the experiment necessarily focuses on long-term memory—anywhere between days and decades, but definitely longer than the hours or even minutes available in a traditional laboratory experiment. We must therefore rely on a different kind of experiment, and on well-known music, memorized before the experiment itself. This is not a challenge in itself—popular music is easy to find. However, even a collection of very well-known popular songs will still contain a lot of music that is unknown to the average participant. To ensure that we can collect enough annotations of each of the items in the collection, a large number of participants is required. The individual listening history of a large group of participants, in turn, may vary widely, so a sufficiently large collection of songs is needed too.

We argue that these requirements—a large number of participants and a large number of songs—make our data collection problem a good candidate for an Internet-based experiment rather than a traditional laboratory experiment.

Serious Games and Games With a Purpose

Research on computational modeling in music tends to rely on music experts to annotate the ground truth data required for model training and evaluation. For some modeling projects, however, experts are unnecessary, or even undesirable (e.g., for music annotations related to mood and emotion or measuring music similarity: annotations for these tasks must reflect the variation in listeners' perception [3]). They require a large amount of annotations from a diverse group of annotators. This makes them good candidates for crowdsourcing. Crowdsourcing—the outsourcing of an automatable task to a virtual crowd of volunteers or paid freelancers—has been a popular and successful strategy for large-scale data collection. Another recent approach to this—a specific kind of crowdsourcing—involves serious games. Serious games are games of which the primary purpose is not to entertain. They have found applications in healthcare, education, and professional training [3].

Serious games used for data collection are also called 'games with a purpose' (GWAP). GWAP and Internet-based experiments are increasingly seen as a serious alternative to lab-based experiments. They can potentially reach a much larger, more varied and intrinsically motivated participant pool, which contributes to the ecological validity of the collected data [69]. GWAP have already proven successful for certain tasks in MIR and machine learning.

For the above reasons, we have decided to frame the experiment as a game. Given that most listeners enjoy catchy music, research on catchiness and hooks seems naturally suited for the GWAP paradigm. With an appropriate choice of music collection we believe the GWAP format allows us to collect enough data for a data-intensive analysis of hooks, and possible future applications in content-based MIR—e.g., hook retrieval or predicting catchiness.

7.3.3 *Gameplay*

The general experiment design followed in Hooked is, therefore, a timed recognition task in the form of a game, in which, for a large number of songs, different sections are presented to a large number of participants.

The hook data collection experiment is a game with four screens as part of the main game loop.

Recognition Screen

When the game starts, a song fragment plays, and the user is asked: "Do you know this song?". This is the *recognition screen*. Participants have a limited amount of time $r_{\max}$ to answer the question; two buttons, YES or NO are shown as in figure 28. To incentivize speed, participants can earn points proportional to the time they have left. The response time is stored together with a song fragment identifier, this is the *recognition time* r.

Figure 28.: Screenshot of the recognition, mute and verification screens in *Hooked!*

Mute and Verification Screens

In a laboratory context, one could expect the participant to answer the above question honestly. In a game context, however, there is an incentive to answer positively regardless of whether the participant knows the song or not. This is why answers are verified in the next two screens.

After a participant hits YES in the recognition screen, the sound of the song fragment will mute for m seconds, and the player is asked to follow along in their head. The *mute screen* is shown, on which the participant can see how much longer the sound will be muted.

After mute time, the sound comes back up, and the song fragment continues playing. However, the fragment might start again from $r + m$ seconds (as if the sound was genuinely muted), or it might start somewhere else inside the song, at time $r + m + \Delta$ (as if the record was scratched)–with equal probabilities. The offset Δ will be referred to as the *distraction* offset. The *verification screen* is then shown, in which the participant is asked whether the music continued in the right place. Again, the response time v (for *verification time*) is recorded.

In other music trivia games, such as *Song Pop*,[4] players are asked to select the correct title or artist of the song to prove they know it. In Hooked, this approach is intentionally avoided: 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 who do know such trivia, the extra cognitive load in recalling it to memory would have an unpredictable and currently irrelevant effect on response time. Instead, the above verification tasks was inspired by the idea that music listeners who know a song well are able to follow along in their heads for some time after the fragment is muted.

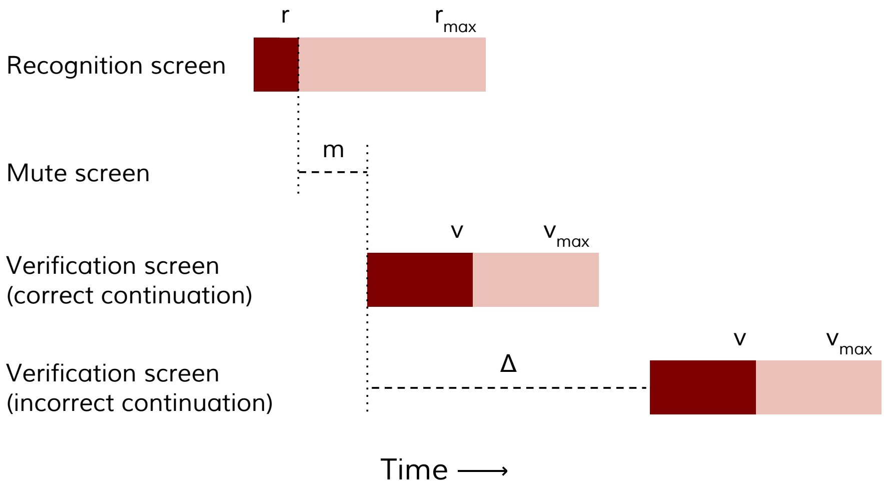
Prediction Screen

Apart from measuring the recognition time as an indicator of ease of recall, we are also interested in participants' intuition on which parts of songs are most catchy. Several times throughout a round of several recognition-verification tasks, Hooked also shows the *prediction screen*. In this screen, participants are asked to listen to two fragments from the same song and choose which they consider catchier—the question reads: 'which fragment is more recognizable'? To give players an extra incentive to give non-random answers, we integrated this task into the game. Each time players complete a prediction task, the chosen fragment is saved. One of the following recognition tasks will then enter a bonus round for double points, presenting the player with the saved fragment that was chosen during the prediction task.

7.3.4 *Experiment Parameters*

The above experiment design involves a number of variables that are recorded and saved to a database as research data: r, v and the answer to all three of the questions (recognition, verification and prediction). Additionally, there are a number of important experiment parameters, including m, $r_{\max}$ and Δ, but also the possible start times of the fragments of the song, which need to be set in advance. Figure 29 shows

4 `http://www.songpop2.com`



Note. The horizontal axis represents time in the song. Vertically separated are three of the four main screens in the game: recognition, mute and verification. In the verification screen, two scenarios are possible: the song continues in the right place, or it continues elsewhere. Audio is playing in the dark red regions.

Figure 29.: Schematic of the parameters and variables in the Hooked experiment.

all variables and parameters in a summary of the time line of the recognition, mute and verification screens of the game.

Start Times

How should start times be chosen? Of course, they could be chosen randomly (e.g., by picking 5 or 7 points in each of the songs in the music collection). However, one of the few agreements in the literature on hooks is that they start at points of considerable structural change [28, 126]. In MIR terms, these would correspond to the boundaries identified by a segmentation system (section 2.2.2). Therefore, start times in Hooked are based on segmentation of the songs by the Echo Nest's structural segmentation algorithm. This not only ensures that starting points are biased to somewhat meaningful locations inside the song, it also ensures that the resulting data can be aligned with other manually and automatically extracted information for the music in The Echo Nest's collections, as they make up an important resource in MIR.

Maximum Recognition Time, Mute Time and Distraction Offset

In the same literature, there is considerably more debate about the *duration* of hooks, something that must be considered when deciding on the right maximum recognition time $r_{\max}$. While songwriters will often speak of the hook as the entire chorus, 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 [101]. The reality very likely sits somewhere in between, but that still leaves a lot of options.

Another challenge is to find the optimal mute time. The time during which participants follow along in their heads shouldn't be so short that one can judge the continuation on the basis of common-sense musical knowledge or timbral characteristics, leading to a type II error. However, it should also not be too long: results from music memory studies have shown that absolute tempo is part of musical memory, though with errors biased towards higher tempi [108]. This would lead to a type I error.

A third parameter that is difficult to establish theoretically is the distraction offset: different offsets might have a different effect on the difficulty of the task and the resulting type I and type II error rate.

To set these three parameters, a pilot experiment was set up, in which 2 options were tested for each parameter. The maximum recognition time $r_{\max}$ was either 10 s or 15 s, the mute time m was either 2 or 4 s, and the distraction offset Δ in the verification task was either 15 s or −15 s. The pilot consisted of a fully functional version of the game, but with just 160 song sections from 32 songs. Twenty-six participants were recruited.

By asking trusted pilot participants to play part of the game honestly, and part of the game competitively, we were able to deduce an estimate of the expected frequency of type I and type II errors. In a statistical model, these were then compared to the different parameter settings. The results, documented in [23], show a trade-off between type I and type II errors, and surprisingly, evidence that honest playing was rewarded with more points, regardless of the parameter settings. Parameter settings were therefore chosen to maximize the pleasantness of playing honestly: $r_{\max} = 15$ s and $\Delta = +15$ s. Mute time did not have a significant effect, so it was left open as subsequent versions of the game were developed.

7.4 IMPLEMENTATIONS

In this section, two implementations of Hooked will be discussed: *Hooked!* and *Hooked on Music*.

7.4.1 Hooked!

The first version of the Hooked experiment was developed for launch in The Netherlands in December 2013. Named *Hooked!*, it was developed for the iOS mobile operating system (iPhone, iPod Touch and iPad). This decision was made after an investigation into the available options for the on line hosting of the copyrighted music fragments. It is difficult to get permission to let users of a game have access to a collection of music for free. Even with these permissions in place, it is difficult to protect on line audio from being used outside the environment of our game. We were able to circumvent these difficulties by working with Spotify's streaming API. Spotify's paid Premium service allows users to stream music to their mobile devices. The streaming API allows developers to use this service inside their own applications, for users who subscribe to the service. Since integration of this streaming API into iOS applications was much more straightforward than integration under other operating systems, *Hooked!* was only im-

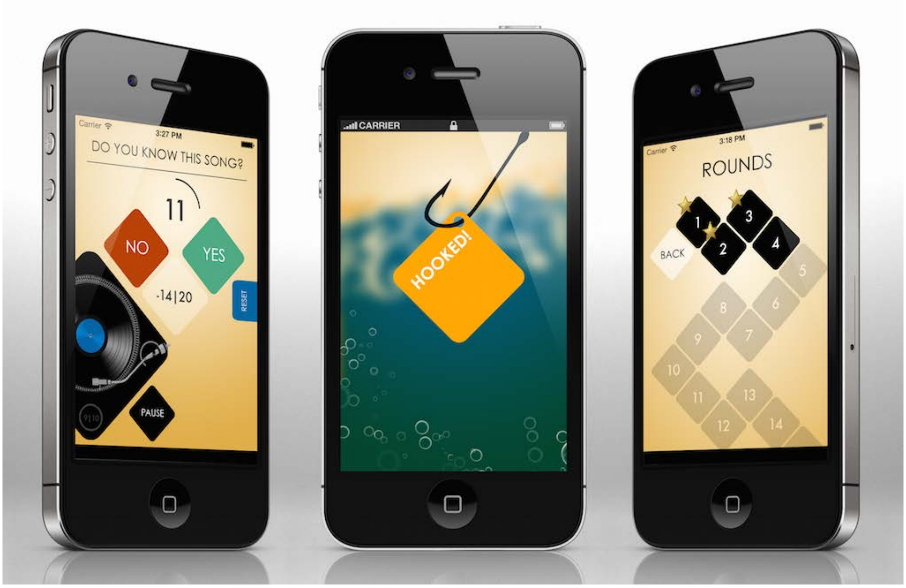


Figure 30.: Visual appearance of the *Hooked!* game. Showing the recognition screen and two more navigation screens (welcome and round selection).

plemented for iOS—realizing that the combined requirements of iOS and Spotify Premium would be a bottleneck for many.[5]

Design

A number of information design aspects of the *Hooked!* game design were completed with advice from Frontwise, an information design agency based in Utrecht, The Netherlands.[6]

A mock-up of the final appearance of the game is shown in figure 30. The recognition, mute and verification screen are shown in more detail in figure 28. In general, the goal in the design process was

5 At the time, however, the market share of iOS was still a lot closer to 50%. Spotify Premium coverage was much lower, but heavily biased to our audience: avid music fans.
6 `https://www.frontwise.com/`

to make the use of *Hooked!* as easy and intuitive as possible. Some of the details that were considered were the color and order of the buttons. E.g.: YES, in the recognition screen, makes the game proceed to a next screen, while NO makes it go back to the recognition screen (but with a different song). Exploiting implicit associations between right-forward and left-backward, the YES button was placed on the right. Were the presentation of instructions (at the beginning of the game) and help text (optionally shown along the way) were also given extensive consideration.

Data

To ensure that the songs used in the game were well-known, the music collection was settled on a subset of the 2012 edition of 'Top 2000', a list of the 'greatest songs of all time' as voted by the listeners of a popular annual radio programme in The Netherlands.[7] As the show is one of the Netherlands' most popular music events of the year, the list represents the musical preferences of a substantial part of the country, including both older and younger generations. The Top 2000 has been characterized as an informal canon of popular music for The Netherlands, though documenting a consensus on popularity rather than quality as judged by experts or the establishment [1].

The advantage of this choice of data is that—unlike with a sample of popular music based on historic chart success—we have some guarantee that the music is still known by many today. On the other hand, any popularity measure based on lists like the Top 2000 has serious drawbacks, too, including somewhat of a demographic and geographic bias, resulting in the absence of some more recent popular styles (e.g., hip hop and electronic dance music, otherwise two very popular genres in the Netherlands). Note, also, that the Dutch audience's preferences in popular music differ from those of other markets (not just because of the presence of music by local artists, but also in the appreciation of particular genres, like hip hop or country

7 `http://www.radio2.nl/top2000`

music). *Hooked!* is therefore suitable only to support models of 'Dutch hooks'—hooks as heard by Dutch popular music listeners.

The *Hooked!* subset of the Top 2000 list included all songs that could be streamed on Spotify, 1591 in total. These songs were divided into 20 groups based on popularity, as measured by The Echo Nest's 'hotttnesss' metric.[8] Each of these groups is used as a level in the game, such that users who finish the game begin (level 1) with the most popular songs and end (level 20) with the lesser-known ones.[9] The resulting unequal distribution of annotations over songs allowed us to collect a sizable number of annotations per song right from the beginning, with more groups of songs collecting annotations as more users joined the game.

Results

After two years, *Hooked!* has been played by 1986 unique players, gathering a total of 167,704 responses to the recognition question. On average, players answered the recognition question positively in 61.5% of the trials, 38.5% of trials were answered negatively or skipped. After a positive answer, the verification question was answered correctly 74% of the time (45.5% of total trials). The top annotated song—Adele's Someone Like You (2011)—was correctly identified 483 times.

7.4.2 Hooked on Music

Building on the experiences developing and publishing *Hooked!*, the COGITCH project partnered with the Manchester Science Festival to produce a more widely accessible version of Hooked, to be launched in 2014. The result, *Hooked on Music*, was a complete reimplementation of the game. It was built primarily in HTML5, with a responsive design that renders on any browser, desktop or mobile, and adapts

8 `http://developer.echonest.com/docs/v4/song.html`

9 It should be noted that the use of the (US-based) Echo Nest popularity data creates a concentration of English language songs at the beginning and Dutch language songs at the end. As a result, many Dutch language songs did not collect enough annotations for further research.

to the size of the screen. Rather than relying on Spotify's streaming API, a music license was negotiated with Dutch collecting societies Buma/Stemra[10] (representing composers and music publishers) and SENA[11] (representing musicians and producers) so that anyone could access the game. Audio was hosted by Soundcloud.[12]

Unlike in *Hooked!*, there are no levels in *Hooked on Music*: for as long as they keep playing, participants are given fragments to recognize, in random order. It also drops the prediction screen in which participants would be asked to predict which of two fragments of a song is more recognizable. The implementation did include a questionnaire through which participants could give us more information about their background (incl. age, level of musical education, estimated weekly hours of music listening). There is also one difference in experiment settings: the mute time (section 7.3.3) in *Hooked on Music* is $m = 4$ seconds. The music used in *Hooked on Music* is a sample of British popular music: 1000 45-second excerpts from 200 chart-topping popular songs from the 1940s to the present (5 excerpts per song).

Over 160,000 participants from 200 countries played *Hooked on Music*, spanning an age range from 15 to 85. A total of over 3 million responses to the recognition question were collected, an order of magnitude more than in *Hooked!*.

7.5 CONCLUSION

In this chapter, we introduced the Hooked experiment on popular music and memory. We presented our motivations to create a dataset of hooks, and reviewed current musicological and cognitive perspectives on the related topics of hooks, distinctiveness and repetition. We then discussed experiment design, explaining our decision to cast the experiment as a game, and two implementations: *Hooked!* and *Hooked on Music*.

10 http://www.bumastemra.nl/en/
11 http://www.sena.nl/en/
12 http://soundcloud.com/

The analysis of the participant data (response times and accuracies) and the audio used in *Hooked!* will be presented in the next chapter. The first iteration of data collection in the *Hooked on Music* game, described above, has only just been closed. The data and audio for this version of the experiment will be analyzed in the months to come (see section 9.3 on current and future work).