Modelling chills in music

Rémi de Fleurian

Submitted in partial fulfilment of the requirements
of the Degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary University of London

February 2022

Declaration

I, Rémi de Fleurian, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below. I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: **Rémi de Fleurian**
Date: **25 February 2022**

Publications

Chapters 2 and 4 are respectively based on the following journal articles:

- de Fleurian, R., & Pearce, M. T. (2021). Chills in music: A systematic review. *Psychological Bulletin*, *147*(9), 890–920. https://doi.org/10.1037/bul0000341
- de Fleurian, R., & Pearce, M. T. (2021). The relationship between valence and chills in music: A corpus analysis. *i-Perception*, *12*(4), 1–11. https://doi.org/10.1177/20416695211024680

Chapters 3 and 5 are currently being adapted for the following submissions:

- de Fleurian, R., & Pearce, M. T. (2022). *Longitudinal study of chills in music: Effects of musical content, stylistic preference, and familiarity.* Manuscript in preparation.
- de Fleurian, R., Benetos, E., Clemente, A., & Pearce, M. T. (2022). *Automatic detection of chills in music.* Manuscript in preparation.

The data collected in Chapters 2 and 5 was released in two publicly available datasets:

- de Fleurian, R., & Pearce, M. T. (2021). *Chills in Music (ChiM).* OSF. https://doi.org/10.17605/osf.io/uyg7m
- de Fleurian, R., & Pearce, M. T. (2021). *Onsets of Chills in Music (oChiM).* OSF. https://doi.org/10.17605/osf.io/x59fm

Early versions of the work presented in this thesis appeared in the following conference proceedings and programmes:

- de Fleurian, R., & Pearce, M. T. (2019). Musical chills: Effects of stimulus properties, stylistic preference and familiarity. In P. Martens, & F. Upham (Eds.), *Abstracts of the 2019 biennial meeting of the Society for Music Perception and Cognition, SMPC 2019* (p. 37).

- de Fleurian, R., & Pearce, M. T. (2019). Effects of stimulus properties, stylistic preference and familiarity on musical chills. In *Programme of the SEMPRE Graduate Conference 2019* (p. 26).
- de Fleurian, R., & Pearce, M. T. (2018). Musical chills: Stimulus properties, stylistic preference and familiarity. In P. Kudumakis, & S. Dixon (Eds.), *Proceedings of the Digital Music Research Network workshop, DMRN+13* (p. 19).

Finally, the following article was also produced during the doctoral study period:

- de Fleurian, R., Harrison, P. M. C., Pearce, M. T., & Quiroga-Martinez, D. R. (2019). Reward prediction tells us less than expected about musical pleasure. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(42), 20813–20814. https://doi.org/10.1073/pnas.1913244116

Abstract

Chills are a fleeting, pleasurable bodily sensation, sometimes accompanied by piloerection, experienced when listening to specific musical passages. They are often used as an indicator of aesthetic and emotional responses to music, because they are considered to be pleasurable, widespread, memorable, and observable. However, research on chills suffers from theoretical and practical limitations. Notably, there are significant shortcomings in the available literature regarding research design, adequacy of experimental variables, and empirical measures of chills. This thesis reviews the suitability of using chills in musical aesthetics and emotion research, and details the construction of several large-scale datasets of pieces of music that elicit chills, in order to solve ongoing issues of reproducibility, generalisability, and ecological validity. These datasets are used in a range of behavioural and computational studies, applying paradigms that address current gaps in research on chills, and particularly the widely hypothesised role of musical expectation and associated factors, which lacks any form of direct empirical verification. More specifically, this thesis presents 1) a systematic review of the literature on chills in music, 2) a longitudinal study using a combination of self-reported and objective measures of chills to assess the roles of musical content, stylistic preference, and familiarity, 3) a corpus analysis investigating conflicting effects of perceived valence on chills, and 4) a large-scale computational study modelling the onsets of chills from acoustic and syntactic properties, resulting in a system with the potential to predict if and when chills might occur in a piece of music. While chills have often been described as idiosyncratic, this thesis demonstrates strong associations between chills and expectation, providing further clarity on competing psychological theories about the origins of chills, and of music appreciation in general.

Contents

List of figures

List of tables

Acknowledgements

I am deeply grateful to my supervisor, Marcus Pearce, for his unwavering support and patience, and for always making the time to provide thoughtful advice. I am also thankful for the endless guidance offered by Emmanouil Benetos, and for the valuable feedback from Bob Sturm, Elaine Chew, and Mathieu Barthet, the other members of my progression panel, and from Diana Omigie and Hauke Egermann, my external examiners. I am indebted to Peter Harrison and Ana Clemente for their help and suggestions, and to the other members of the Music Cognition Lab for promoting high standards of critical thinking and scientific integrity, and providing countless opportunities for stimulating discussion. I would like to acknowledge the support offered by the Cognitive Science Research Group and the Centre for Digital Music, as well as the Media and Arts Technology programme and its staff—Jonathan Winfield in particular, who facilitated much of my work. I am also grateful to the open-source communities behind the multiple software tools on which the present research depended. Finally, I would like to thank Hannah for being profoundly encouraging and kind throughout my doctoral studies.

This research was supported by the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology [EP/L01632X/1].

Chapter 1

Introduction

1.1 Empirical aesthetics

Music is a human universal (Mehr et al., 2019; Savage et al., 2015) and is one of the most commonly reported sources of emotional pleasure (Dubé & Le Bel, 2003). Yet, despite the prevalence of musical behaviours across cultures, the nature of the relationship between music, pleasure, and emotion is poorly understood.

Empirical aesthetics were one of the very first topics of interest in experimental psychology. Early work by Gustav Fechner, Hermann von Helmholtz, and Wilhelm Wundt looked into the effect of stimulus complexity (Wundt, 1863), relationships of musical tones (Helmholtz, 1863), and learned associations (Fechner, 1876) on aesthetic experience, and inspired a long tradition of psychological research into what elicits pleasurable responses to art (for a review, see Huron, 2016).

One notable contribution to the field comes from Berlyne (1971), who further investigated the relationship between complexity, arousal and pleasure, leading to the hypothesis that there exists an inverted-U relationship between hedonic value and arousal potential, in which arousal is influenced by stimulus properties such as complexity and familiarity, and pleasure is higher for moderate degrees of arousal, and lower for both high and low degrees of arousal. This hypothesis has been extensively tested in various artistic domains, including music, for which it has received relatively strong (e.g Heyduk, 1975; North & Hargreaves, 1995; Vitz, 1966), though not unanimous (e.g. Orr & Ohlsson, 2005; Smith & Melara, 1990) empirical support (for reviews, see Chmiel & Schubert, 2017; Hargreaves & North, 2010; Orr & Ohlsson, 2005).

The methods used in such research have become problematic in a few ways. First, they distinguish aesthetic judgement from emotional response (Orr &

Ohlsson, 2005) and place the focus on preference, when more comprehensive accounts suggest a complex relationship between aesthetic response, emotion, and liking (Juslin et al., 2010), involving social, emotional, and cognitive components (Konečni, 1979), and depending on a reciprocal interplay between listener, context, and music (Gabrielsson, 2001a; Hargreaves, 2012). As a result, while a focus on the relationship between preference and stimulus properties might enable convenient experimental approaches, it is unlikely to reflect the broad nature of affective and aesthetic responses and the psychological mechanisms that underlie them (Huron, 2016; Juslin, 2016).

Second, research designs in empirical aesthetics have historically relied on tightly controlled experimental procedures in laboratory environments. While there are many benefits to this approach, it encourages the use of stimuli which only approximate music, in short-lived situations which are unlikely to induce fully fledged emotional and aesthetic responses. Ecological validity often comes at a cost, but when it comes to empirical aesthetics, it has been recommended to make use of more naturalistic listening experiences (Hargreaves & North, 2010; Hodges, 2016), in addition to considering the possibility of employing longitudinal designs to study the development of responses to these experiences over time (Greasley & Lamont, 2016).

Third, Berlyne's (1971) and subsequent approaches do not always emphasise the temporal nature of music, and tend to consider musical stimuli as a whole instead. Music being a phenomenon that unfolds in time, dynamic fluctuations in affective and aesthetic responses should be taken into consideration (Huron & Margulis, 2010; McDermott, 2012). Madsen et al. (1993), for instance, used continuous self-report methods, and found that the trajectories of aesthetic ratings are highly consistent within a piece of music for individual listeners, highlighting the need for the study of the underlying factors at the origin of these consistent patterns.

Lastly, there remain substantial challenges because of the lack of objective experimental variables with the potential to capture subjective experiences of pleasure and emotion. Self-report measures and physiological responses are traditionally used, but both have disadvantages in terms of susceptibility to biases, reliability, and specificity (Juslin, 2016; Larsen et al., 2008; Orne, 1962; Panksepp & Bernatzky, 2002; Zentner & Eerola, 2010).

1.2 Musical expectation

Aesthetic experience depends on many factors. One prominent psychological phenomenon thought to be particularly relevant to empirical aesthetics is musical expectation. Research on the topic benefits from a long-standing theoretical background initiated by Eduard Hanslick (1854), and reinforced by Leonard Meyer (1956, 1957) in parallel to the research discussed above on complexity, arousal, and pleasure.

Musical expectation is based on the hypothesis that developing expectations follows a process of probabilistic learning of the statistical regularities in musical structure (Pearce, 2018; Saffran et al., 1999). In other words, with exposure to a musical culture, listeners automatically and implicitly develop an internal model of the structure of a musical style through a process called statistical learning, which is then used, when listening to music, to form expectations about the possible continuations of the music through a process called probabilistic prediction (Pearce, 2018). These learned expectations can be violated, delayed, or confirmed, resulting in induced emotional and aesthetic responses (Cheung et al., 2019; Egermann et al., 2013; Gold et al., 2019; Huron, 2006; Juslin, 2013; Sauvé et al., 2018; Steinbeis et al., 2006), possibly in order to drive learning to improve future predictions or otherwise optimise states of arousal (Pearce, 2018).

Much research has been conducted on the topic of musical expectation. Behavioural methods have been used to demonstrate effects of melodic, harmonic, rhythmic or metrical expectation on recognition memory, music production, music perception, and music transcription (for a review, see Pearce & Wiggins, 2012). Research using neuroscientific methods has shown that violations of expectations involve activity in inferior frontal regions, the caudate, and the nucleus (for reviews, see Salimpoor et al., 2015; Trainor & Zatorre, 2016). Finally, the use of computational methods has revealed that several modelling approaches can be used to successfully predict expectations (for a review, see Rohrmeier & Koelsch, 2012).

Naturally, some of the methodological limitations identified with regards to empirical aesthetics also apply to research on the relationship between expectation, emotion, and pleasure, which generally consists of gathering self-reports in response to relatively short (Cheung et al., 2019) and often manipulated (Gold et al., 2019; Sauvé et al., 2018; Steinbeis et al., 2006) melodic and harmonic sequences.

1.3 Music-evoked chills

One particular reaction to music, commonly referred to as chills, appears related to both aesthetics and expectation. There is little agreement on the exact definition of chills, on their physiological basis, or on their relationship to emotions. This motivated a systematic review of the literature on chills, presented in Chapter 2, based on which we define chills for the purpose of the present thesis as a fleeting, pleasurable bodily sensation, sometimes accompanied by goosebumps, experienced when listening to specific musical passages.

Music-evoked chills (MECs) are often considered to be a pleasurable response, and are thought to be stable, memorable, discrete, and observable, therefore addressing some of the limitations of self-reports and physiological measures. As a result, they have emerged as a convenient indicator of emotional and aesthetic experiences in research on responses to music. However, there is no consensus on how MECs relate to emotion, aesthetics, and pleasure, which motivates the need for further study. In fact, as revealed in Chapter 2, we suspect that MECs are an optional but enhancing component of aesthetic and emotional responses, which therefore makes them unsuitable as the sole indicator of such responses.

Musical expectation has long been posited as a cause of MECs (L. Harrison & Loui, 2014; Huron, 2006; Huron & Margulis, 2010; Juslin, 2013; Juslin & Västfjäll, 2008; McDermott, 2012; Mencke et al., 2019; Pearce & Wiggins, 2012; Salimpoor et al., 2011; Sloboda, 1991), but empirical investigations of this relationship suffer from theoretical and practical limitations, as does much of the wider research on MECs. Notably, there are significant shortcomings in the available literature regarding research design, adequacy of experimental variables, and empirical measures of MECs.

1.4 Thesis overview

The exact nature of the relationships between MECs, expectation, and affective and aesthetic responses remains unclear. The study of such relationships should aim to address some of the limitations which have historically affected research in empirical aesthetics, and stand to gain from the adoption of typically underused methodological approaches. Notably, computational methods applied to large collections of naturalistic stimuli are particularly well suited to the study of MECs, but have yet to be used, despite the success of similar approaches in research on music and emotion (e.g., Eerola, 2011).

Specifically, research on MECs should address the limitations discussed above by providing clarity on the relationship between MECs, aesthetic, and emotional responses, making use of naturalistic listening situations, and considering musical events as they dynamically unfold in time rather than as an aggregated whole. The present thesis seeks to do so by reviewing the suitability of using MECs in musical aesthetics and emotion research, detailing the construction of several large-scale, naturalistic datasets of pieces of music that elicit MECs, and using such datasets in a range of behavioural and computational studies. The presented research applies paradigms that address current gaps in the available knowledge on MECs, and particularly the widely hypothesised role of musical expectation and associated factors, while attempting to solve ongoing issues of reproducibility, generalisability, and ecological validity.

The context for the present work is presented as a systematic review of MECs in Chapter 2. A longitudinal study of MECs is discussed in Chapter 3, exploring the effects of musical content, stylistic preference, and familiarity. Computational studies of MECs are detailed in Chapters 4 and 5, investigating expressed valence and musical expectation respectively. The thesis concludes with a discussion of the outcomes and implications of the present work in Chapter 6.

Chapter 2

Systematic review

With the literature doubling in size since the last review on MECs, it has become increasingly difficult to gain a broad and integrated understanding of the empirical and theoretical research on the subject. Notably, crucial questions remain about the criteria that are necessary and sufficient to characterise MECs. In this chapter, we systematically review the literature on MECs in order to reconcile diverging opinions and empirical findings on their psychological nature, and to develop a preliminary model that provides a robust framework for future hypothesis-driven research. We explore the context behind current research on MECs, discuss how they relate to emotional and aesthetic responses, assess current empirical measures and paradigms, summarise their physiological and neural correlates, categorise their possible stimulus-driven elicitors, examine how they are affected by individual differences, and evaluate theoretical perspectives about their potential evolutionary causes. We conclude by providing a preliminary model of MECs that suggests different pathways for the experience of MECs, a dataset listing pieces of music reported to elicit MECs in the reviewed literature, and a set of open issues, hypotheses, and recommended approaches for future research.

2.1 Introduction

The knowledge base on MECs is rapidly expanding, and as research findings accumulate, it is becoming increasingly difficult to gain a comprehensive and integrated psychological picture of what MECs entail. Notably, crucial questions remain about the criteria that are necessary and sufficient to characterise MECs.

Frequently cited papers describe MECs as "a spreading gooseflesh, hair-on-end feeling that is common on the back of the neck and head and often moves down the spine" (Panksepp, 1995, p. 173), "a particularly intense, euphoric response to music [frequently accompanied] by an autonomic or psychophysiological component" (Blood & Zatorre, 2001, p. 11818), "intense emotional experiences involving sensations such as goose bumps or shivers down the spine" (Koelsch, 2010, p. 131), or "a pleasant tingling feeling associated with the flexing of hair follicles, resulting in gooseflesh (technically called piloerection) accompanied by a cold sensation, and sometimes producing a shiver" (Huron & Margulis, 2010, p. 591).

While superficially similar, these definitions provide pointers to crucial questions which need to be addressed. If MECs are to be used as an indicator of pleasurable experiences, it is important to understand how universal and frequent they are, as well as the nature of their relationship with emotional and aesthetic responses, in order to assess whether or not their relevance is justified, and, if so, clarify their underlying psychological mechanisms. The phenomenology of MECs also deserves clarification, as it is unclear whether empirical findings refer to a single psychophysiological response or to distinct experiences with little common ground. The specificity of the physiological and neural signatures of MECs needs to be explored to establish whether MECs invoke general-purpose mechanisms involved in other functions, such as emotional processing and reward, or are distinguishable from these experiences. Finally, it is necessary to investigate the causes of MECs, both in terms of stimulus-driven properties and individual differences, to better understand their origin, and thereby achieve a broader and more integrated understanding of the empirical and theoretical research on MECs.

MECs are often mentioned in the literature on music and emotion, but at the time of writing, there are only three short reviews entirely dedicated to MECs (Grewe et al., 2009b; L. Harrison & Loui, 2014; Mori & Iwanaga, 2014a), one review about MECs and the autonomous sensory meridian response (del Campo & Kehle, 2016), one review about MECs and music therapy (Tihanyi, 2016), one philosophical essay about MECs and musical aesthetics (Levinson, 2006), two book chapters discussing MECs within the context of musical expectation (Huron & Margulis, 2010) and of the evolutionary basis of music (Altenmüller et al., 2013), and book chapters on music and emotion which contain subsections on MECs (e.g., Corrigall & Schellenberg, 2013, 2015; Hodges, 2016; Hunter & Schellenberg, 2010; Juslin, 2019; McDermott, 2012; Sachs et al., 2018; Stark et al., 2018; Vuust & Kringelbach, 2010).

Despite referring to the same phenomenon, as evidenced by the fact that these contributions all make reference the same seminal papers on MECs (Blood & Zatorre, 2001; Goldstein, 1980; Panksepp, 1995; Sloboda, 1991), the topics listed above are very diverse, once again illustrating the need for a clear integration of the 40 years of available research on MECs. The purpose of this chapter is therefore to systematically review the literature on MECs in order to reconcile diverging opinions and empirical findings on their psychological nature, and to develop a preliminary model that provides a robust framework for future hypothesis-driven research.

2.2 Methods

We performed a systematic literature search in order to ensure comprehensive coverage. We opted to conduct a systematic review instead of a meta-analysis because of the great diversity of topics and methods in research on MECs, which results in insufficiently comparable research evidence for a quantitative aggregation of empirical findings. We first outline the search procedure, before going over the inclusion and exclusion criteria, and finally describing how the findings are organised in the next sections of the present chapter.

2.2.1 Literature search

We searched the databases Web of Science, APA PsycInfo and PsycExtra, PubMed, Scopus, and Google Scholar, using a cut-off date of 30 April 2020, for articles, reviews, conference papers, books, book chapters, and doctoral dissertations about chills and music. All contributions containing the term *music* and at least one of *chills*, *thrills*, *frisson*, *shivers*, *goosebumps*, or *piloerection* were considered, resulting in 149 records being identified on Web of Science, 85 records on APA PsycInfo and PsycExtra, 47 records on PubMed, and 127 records on Scopus. We also examined the first 100 records returned by Google Scholar for the same search terms, as well as the first 100 records on Google Scholar for contributions dated 2019 or later to ensure we did not miss recent contributions. This process resulted in the identification of 346 unique records.

2.2.2 Inclusion and exclusion criteria

The objective was to include all publications about MECs. We therefore included contributions written in any language, as long as they mentioned both chills and music. The first exclusion criterion accounted for the fact that the queried terms are commonly used in the English language, and therefore appear in many publications which are not about MECs. As a result, 117 irrelevant records were excluded. The second exclusion criterion accounted for the fact that MECs are often briefly mentioned to provide context in broader studies, reviews, or book chapters about music and emotion. As a result, 78 records containing no substantial information about MECs were excluded. In addition, we excluded five records that could not be retrieved, one article written in Japanese that could

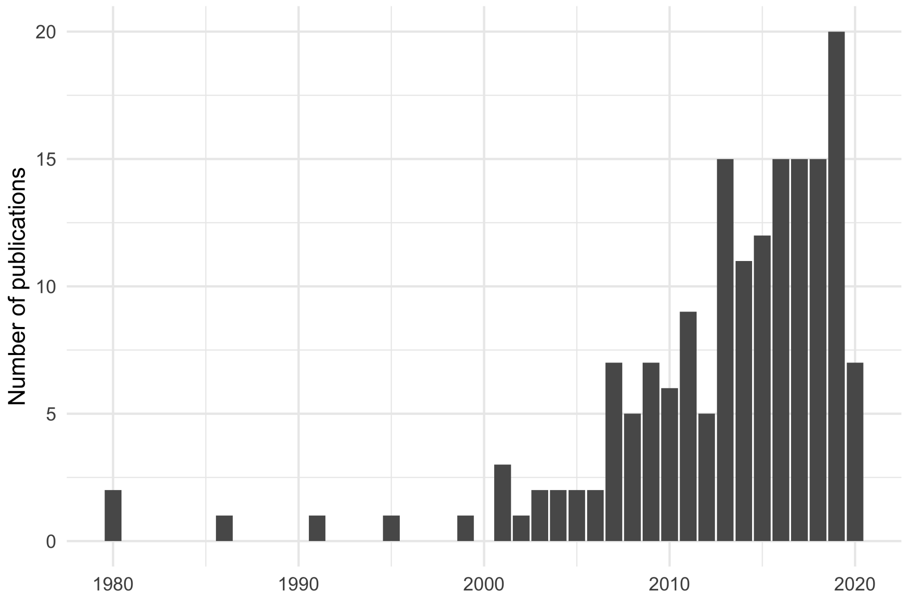

Figure 2.1: Yearly publication count for research on the topic of MECs reviewed in this chapter.

not be translated online due to issues with character encoding, one corrigendum, the content of which was already reflected in the associated publication, one editorial which simply listed the topics covered in a specific journal issue, and six records because the presented results were also fully covered in subsequent journal articles that were retained in the search. Finally, we included 30 articles and book chapters obtained through backward and forward reference searching, resulting in a total of 167 contributions which represented, to our knowledge, all the available academic literature on MECs.

2.2.3 Organisation of findings

The literature doubled in size since the reviews by L. Harrison and Loui (2014) and Mori and Iwanaga (2014a)—in the present chapter, we reviewed 83 contributions about MECs dated 2014 or prior, and 84 contributions dated 2015 or later (see Figure 2.1 for the yearly publication count).

The vast majority of publications on MECs contain findings that pertain to several domains of interest, which logically emerged as empirical and theoretical findings were extracted from each reviewed study. As a consequence, instead of attempting to allocate the publications themselves to meaningful units, we distributed all of their findings across several overarching categories corresponding to these domains, therefore allowing broad and integrative coverage of the most pertinent and widely researched topics in research on MECs. The results are therefore structured as follows. Sections 2.3.1, 2.3.2, and 2.3.3 consider the wider context within which empirical and theoretical research on MECs has been conducted. We begin in Section 2.3.1 by considering terminological issues, the phenomenological nature of MECs, their prevalence and frequency, and their relationship with other psychological processes. In Section 2.3.2, we expand on the nature of the relationship between MECs, pleasure, and emotional and aesthetic experience, before assessing subjective and objective ways of measuring MECs, as well as experimental paradigms used in research on MECs in Section 2.3.3. In the subsequent sections of the chapter, we review the empirical literature on the biological basis of MECs, considering associations between MECs, arousal, and physiological responses (Section 2.3.4), and neural correlates of MECs in the basal ganglia and other brain structures (Section 2.3.5). We then turn to theoretical considerations regarding the causes of MECs. We review the empirical literature to identify the stimulus-driven causes of MECs and categorise them into acoustic, musical, and emotional elicitors (Section 2.3.6), examine empirical effects of individual and personality differences on the occurrence of MECs (Section 2.3.7), and critically evaluate the degree of support provided by the reviewed evidence for current theories on the function of MECs (Section 2.3.8). These findings are summarised and expanded upon in the discussion, and the quality of the reviewed research is evaluated, after which we conclude by providing a preliminary model of MECs, a dataset listing pieces of music reported to elicit MECs in the reviewed literature, and a set of hypotheses and recommendations for future research.

2.3 Results

2.3.1 Context

A significant amount of research has focused on identifying exactly what MECs are, but there remains uncertainty about many of their defining aspects. In this section, we review the terminology associated with MECs, their phenomenological nature, their prevalence and frequency, and their relationship with other psychological processes, including emotional and aesthetic responses to music.

Terminology

Besides definitions of MECs, an initial source of confusion is the broad range of terms used to refer to the phenomenon. Terms such as musical chills, aesthetic chills, art-elicited chills, shivers, shivers down the spine, psychogenic shivering, thrills, frisson, goosebumps, gooseflesh, goose pimples, piloerection, emotional piloerection, hair standing on end, and skin orgasm, have been used interchangeably over the years, and there is no explicit consensus as to which option should be preferred. L. Harrison and Loui (2014) recommended the use of *frisson*, a term first used in the context of research on MECs by Huron (2006) and Levinson (2006), which has the advantage of providing a relatively nonspecific way to describe an emotional response with a physiological component, while avoiding the burden of cultural associations present in other terms. While this is a sound recommendation, the term *frisson* is sparsely used in the literature. We would argue that the need for a unified term of reference outweighs considerations about the colloquial use of the term, and therefore recommend the use of *chills*[1], which has quite clearly become the most prevalent term in the recent literature. In the present thesis, we use *chills* (for the psychophysiological response) and *piloerection* (for goosebumps specifically) throughout, except when referring to specific findings from authors who used several terms in a single publication.

Phenomenology

Regardless of the terminology used, it is important to have a clear and consistent conception of the nature of MECs. This would ensure that participants in research on MECs provide responses about the same psychophysiological phenomenon. Failing to do so might lead to inconsistent empirical findings, making interpretation problematic and creating difficulties in relating empirical results between studies. However, identifying a clear and consistent phenomenological description of MECs is not straightforward in the existing literature. Goldstein (1980) provided a thorough starting point through a series of unstructured and structured questionnaires, in which several groups of participants were asked to describe their experience of MECs. The results characterised MECs as a transient, pleasurable response associated with sudden changes in mood or emotion, commonly experienced by a large proportion of the population, and originating primarily in the upper spine or back of the neck, with other common points of origin being shoulders, lower spine, and scalp. Intense occurrences of MECs were described as longer in duration, and radiating to other body areas (most commonly the scalp, arms, shoulders, spine, and face). There are further, varying reports of the location from which MECs originate. The back (or spine), head (or scalp, face, or neck), and arms are the most commonly reported points of origin (Craig, 2005; Goldstein, 1980; Neidlinger et al., 2017; Panksepp, 1995; Wassiliwizky et al., 2015), with occasional mentions of hands or fingers (Craig,

[1]Following common usage in the literature, we use the word "chills" as a plural-only, non-countable noun, like clothes or groceries. We feel this is consistent with the difficulty of identifying exactly what would constitute an individual chill (or a definite number of chills) and find it more natural to refer, for example, to an episode of chills.

2005), as well as legs (Wassiliwizky et al., 2015).

Interestingly, Craig (2005) made the distinction between points of origin for shivers or tingling (listed above) and piloerection, which was most often reported to begin on the arms, back of the neck, or legs. This raises the important question of whether piloerection should be considered as an integral component of MECs or not. Again, opinions differ. While some definitions of MECs suggest that piloerection is required (Huron & Margulis, 2010; Panksepp, 1995), most do not (e.g., Blood & Zatorre, 2001; Goldstein, 1980), and empirical findings support the latter view. In self-reports, piloerection is often reported to happen less often than MECs (Gabrielsson, 2011; Silvia & Nusbaum, 2011; Sloboda, 1991). In experimental settings, piloerection was only observed in 57% (Craig, 2005), 40% (Benedek & Kaernbach, 2011), 43.1% (Sumpf et al., 2015), and 40.7% (Wassiliwizky, Koelsch, et al., 2017) of participants who reported MECs. Seemingly, not all MECs involve piloerection (Craig, 2005), although most (Benedek & Kaernbach, 2011) or all (Craig, 2005) occurrences of piloerection were found to happen during experiences of MECs. It is therefore likely that MECs, as reported by participants, do not always involve piloerection, although it is possible that experienced MECs might require an intensity threshold to be reached before piloerection can be observed (Sumpf et al., 2015), or that current piloerection detection methods are simply not accurate enough (for an overview of available methods, see Section 2.3.3). While relying on self-reported or observed piloerection to study MECs is tempting, due to the objectivity it provides, it seems more appropriate at this stage to combine such an approach with self-reports of MECs (e.g., Wassiliwizky, Koelsch, et al., 2017), in order to avoid biasing research away from what people actually experience as MECs (Maruskin et al., 2012).

Prevalence and frequency

While 79% of the 249 participants who completed Goldstein's (1980) questionnaires reported having experienced MECs in the past, additional figures about the prevalence of the ability to experience MECs are available in the literature: 90% of a sample of 83 respondents for experiencing shivers down the spine at least once in the past five years (Sloboda, 1991, also reporting 62% for goose pimples and 31% for trembling), over 80% of 186 respondents for experiencing shivers down the spine or goose pimples at least rarely over the past five years (Mlejnek, 2013), or 86% of 828 respondents for experiencing MECs with some regularity (Panksepp, 1995). In a survey of 196 people by Nusbaum and Silvia (2011), 8% of respondents never or rarely experienced MECs, and in a survey of 188 people by Silvia and Nusbaum (2011), 11.2%, 9.6%, and 23.5% never or rarely experienced chills down the spine, goosebumps, and feeling hair standing on end, respectively, although it is worth keeping in mind that for the latter study, only half of the reports were about experiences when listening to music. There are further figures available in the literature, showing MECs as generally less prevalent in experimental settings (e.g., Colver & El-Alayli, 2016; Grewe et al., 2009a; Konečni et al., 2007), but when looking at prevalence, it makes sense to consider only results from surveys of a reasonably representative sample of the population, since participants in lab experiments have most often been recruited for their ability to experience MECs, but might also not have been able to experience MECs under experimental conditions for a variety of reasons. Limitations remain, due to the fact that people interested in taking surveys about reactions to music might not be fully representative of the population, but from these results, it is reasonable to assume that 90% is an upper limit for the proportion of the population that has the ability to experience MECs. Interestingly, when providing free reports of their strongest, most intense experience of music, respondents spontaneously included MECs or shivers in 10% of their reports, and piloerection or gooseflesh in 5% of their reports (Gabrielsson, 2011).

In terms of frequency, those who experience MECs seem to do so quite regularly. MECs are reported as the most frequent (Sloboda, 1991) or second most frequent physical response to music, behind tears (Gabrielsson, 2011; Scherer et al., 2001), and happen with some regularity for most people (Panksepp, 1995), ranging from every week to every few months (Bannister, 2020b; Goldstein, 1980). For instance, during a week of experience sampling, 81% of respondents reported having at least one experience of MECs, and reported MECs in 14% of the occurrences of listening to music (Nusbaum et al., 2014).

Relation to other psychological processes

Another step in better understanding MECs is to examine the role they play in emotional and aesthetic responses to music, with studies in which such responses are classified using content analysis, factor analysis, or principal component analysis. Panzarella (1980) found that MECs belong to one of the four major dimensions which can describe intense, joyous experiences of listening to music or looking at visual art. This dimension, called *motor-sensory ecstasy*, was found to be mostly associated with the climactic stage of an aesthetic experience. Scherer et al. (2001) coded qualitative reports of the last time respondents were emotionally affected by a piece of music, and assigned MECs and piloerection to one of five major emotion components, called *physiological symptoms*. Gabrielsson and Wik (2003), as a part of their work on identifying the components and causes of strong experiences related to music (Gabrielsson, 2001b), found that in descriptions of the strongest, most intense experiences of music reported by almost 900 participants, MECs and piloerection were best coded and classified as *physiological reactions*, a sub-component of *physical reactions and behaviours*. Zentner et al. (2008), in a series of studies aimed at identifying and validating a taxonomy of musically induced emotions for the development of the Geneva Emotional Music Scale, retained MECs as one of 40 items present in a second-order model of musical emotions. MECs were found to belong to one of nine first-order factors, *transcendence*, which itself belongs to one of three second-order factors, *sublimity*. Silvia and Nusbaum (2011) found that out of twelve unusual aesthetic states, the three states related to MECs (chills down the spine, hair standing on end, goosebumps) made up one of three factors, simply called *chills*. The three factors (chills, touched, absorption) all loaded strongly on a single higher-order factor for *aesthetic experience*. In developing the Barcelona Musical Reward Questionnaire, Mas-Herrero et al. (2013) included an item about MECs as one of twenty items that best capture individual differences in how people experience reward associated with music. This item loaded highly on one of five factors, named *emotional evocation*. Bannister (2020b) coded a large number of reports of how surveyed participants felt during the experience of MECs, and identified *emotions and feelings* and *physical reactions* as the two themes accounting for most responses. Finally, Cotter et al. (2018) used MECs as an item in a twenty-four-item questionnaire about feeling like crying in response to music. The two resulting latent classes were named *awe* and *sad*, with higher levels of experiencing MECs for the former than for the latter—a finding that was replicated in a subsequent study (Cotter et al., 2019).

Two contributions using similar approaches deserve particular consideration, due to their exclusive focus on the experience of chills. Maruskin et al. (2012) put forward a convincing argument that chills might consist of a set of distinct phenomena with different psychological and biological bases. This motivated an extensive body of work in which a wide range of self-reports of the experience of chills associated with emotionally significant events were analysed in order to gain a better understanding of chills as a psychological construct. It was found that chills are best understood as comprising four conceptually distinct sensations: goosebumps, tingling (grouped together as a higher order factor, *goosetingles*, associated with positive affective states), coldness, and shivers (grouped together as *coldshivers*, associated with negative affective states). Similarly, Bannister (2019), using a quantitative approach, investigated whether chills should be considered as a single psychological construct, reflective of intense pleasure and emotion, or as an umbrella term for distinct experiences. Analysis of responses to questionnaire items revealed that chills can be conceptualised as comprising three categories: *warm chills* (associated with positively valenced feelings and physical responses), *cold chills* (associated with negatively valenced feelings and physical responses), and *moving chills* (associated with more ambiguous responses, such as tears, feeling a lump in the throat, affection, or tenderness, among others). Although it is tempting to draw parallels between the categories identified by Maruskin et al. (2012) and Bannister (2019), they are not directly comparable because they were derived from responses to emotionally significant events in one case, and aesthetic stimuli in the other. Regardless, these considerations are of particular importance, because if chills are indeed a collection of phenomenologically and psychologically distinct experiences, failing to distinguish between them might lead to null, conflicting, or misleading results (Bannister, 2019; Maruskin et al., 2012). Note, however, that the vast majority of research on MECs continues to treat them as a single construct.

It is worth noting that several studies reviewed in this section and in the rest of this chapter do not exclusively pertain to reactions to music. These studies were included if they counted music as one of several investigated modalities, or if they reported results relevant to research on MECs. For instance, chills are known to occur in response to visual stimuli (Bannister, 2019; Goldstein, 1980; Grewe et al., 2011; Maruskin et al., 2012; Panzarella, 1980; Silvia & Nusbaum, 2011; Sumpf et al., 2015; Wassiliwizky, Jacobsen, et al., 2017), and also to text, poetry, film audio, sounds (human, animal, natural, and technical), speech, beauty in nature, touch, smell, taste, memories, and virtual reality environments, among others (Benedek & Kaernbach, 2011; Bériachvili, 2016; Goldstein, 1980; Grewe et al., 2011; Konečni et al., 2007; Quesnel & Riecke, 2018; Schoeller & Eskinazi, 2019; Schurtz et al., 2012; Wassiliwizky, Koelsch, et al., 2017). In cases where occurrences of chills were compared across modalities, there is no consensus as to whether music should be considered the most potent elicitor (Goldstein, 1980; Sumpf et al., 2015) or not (Bannister, 2019; Benedek & Kaernbach, 2011; Grewe et al., 2011; Schurtz et al., 2012). Two of these studies set out to answer that question explicitly through surveys (Goldstein, 1980; Schurtz et al., 2012), while the other analyses of this effect simply compared occurrences of chills across the specific sets of stimuli used in each study, making it difficult to assess how generalisable these results are.

2.3.2 Emotion and aesthetics

As discussed in the previous section, MECs have been fairly consistently classified as components of emotional or aesthetic experiences. However, there is also considerable discussion about what constitutes such experiences, and therefore their specific relationship with MECs deserves clarification. In this section, we review how MECs are associated with emotional responses, pleasure, and aesthetic responses.

Emotional Response

MECs are often discussed in book chapters on music and emotion, either as a physiological response which can accompany intense musical emotions (Juslin, 2016), or as a strong, specific emotional reaction to music (Eerola, 2018; Hunter & Schellenberg, 2010). To disentangle these interpretations, it is useful to refer to definitions of musical emotions. MECs show some of the qualities of emotional states, as defined by Juslin et al. (2010), because they can involve a subjective experience, observed in self-reports of emotional reactions to music, as discussed earlier, and because they have been shown to involve physiological arousal, both in terms of measured physiological responses and self-reported arousal (see Section 2.3.4). However, MECs do not clearly exhibit other characteristic components of emotional states, such as *motor expression* or *action tendency* (Juslin et al., 2010; Scherer, 2009), and they can be associated with positive or negative valence (e.g., Bannister, 2019; Maruskin et al., 2012). These considerations suggest that, instead of being considered as an emotion category or emotional state per se, MECs are best understood as a psychophysiological response which can form part of a range of emotional states (Grewe et al., 2011; Juslin, 2019).

Pleasure

In this thesis, we make a distinction between pleasure experienced while listening to music and positively valenced music-evoked emotion (see E. Schubert, 2013). It is perfectly possible, for example, to experience sadness while listening to a piece of music but also to find that experience pleasurable. Most studies of MECs have treated them as a pleasurable response to music. Interestingly, this notion permeated the early literature on MECs despite limited evidence at the time that MECs were indeed associated with pleasure (Blood & Zatorre, 2001; Goldstein, 1980). Since then, research has confirmed that such an association exists, as shown by an analysis of qualitative reports in an extensive survey (Bannister, 2020b), by significant increases in pleasure occurring immediately prior to the onset of MECs and peak pleasure coinciding with MECs (Salimpoor et al., 2009), by a joint increase in pleasure and occurrence of chills when watching video clips preceded by a meaningful statement as opposed to an incoherent statement (Schoeller, Eskinazi, & Garreau, 2018; Schoeller & Perlovsky, 2016), by MECs playing a role in driving music preference (Schäfer & Sedlmeier, 2010, 2011), and more generally, by a documented association between MECs and self-reports of increased subjective pleasure when listening to music (Grewe et al., 2011; Grewe et al., 2009a; Grewe et al., 2007; Mori & Iwanaga, 2014b, 2015, 2017;

Salimpoor et al., 2011; Salimpoor et al., 2009; Sumpf et al., 2015). Interestingly, displeasurable chills can also be experienced in response to unpleasant sounds (Grewe et al., 2011; Grunkina et al., 2017; Halpern et al., 1986; Klepzig et al., 2020). Given that chills can form a part of unpleasant experiences, it is possible that MECs are generally experienced as pleasurable because music listening itself is generally a pleasurable activity (Dubé & Le Bel, 2003).

Aesthetic response

Since MECs are generally experienced as pleasurable, their role in aesthetic responses also deserves clarification (Hodges, 2016). MECs have been referred to as one of several indices of aesthetic experiences of music (E. Schubert et al., 2016; Vuust & Kringelbach, 2010). As noted earlier, previous questionnaires and qualitative reports about aesthetic responses to music have included MECs (Panzarella, 1980; Silvia & Nusbaum, 2011). To better understand this relationship, we need a precise definition of the aesthetic appreciation of music. Here, we follow Levinson (2009) in characterising aesthetic appreciation as a positive estimation based on an intrinsically pleasurable experience arising from attention directed to the form and content of a piece of music. Based on the range of psychological components thought to be involved in aesthetic appreciation (see Leder et al., 2004; Leder & Nadal, 2014, for another extensive, multi-component model), it seems unlikely that MECs should be considered as an aesthetic experience in and of themselves. Rather, a more promising interpretation would be that MECs can contribute to aesthetic experiences, because they constitute a pleasurable response to some musical properties (see Section 2.3.6). Indeed, in a philosophical essay about MECs, Levinson (2006) argues that they provide a signal that something significant happened in the music—in other words, a focuser of attention—and in so doing, make a valuable contribution to wholly experiencing a piece of music, through a culmination of cognitive, emotional, physiological, and behavioural responses. According to E. Schubert et al. (2016), this contribution, and that of other subjective experiences evoked by music (or *internal locus* affects), is what motivates people to seek out aesthetic experiences. Many researchers have considered MECs to form an optional, rather than a central, component in the aesthetic experience of music (e.g., Bériachvili, 2016; Brattico et al., 2013; Gabrielsson et al., 2016; Konečni, 2007), and this is a view we share, in light of the reviewed literature.

2.3.3 Measures and paradigms

Most of the early research on MECs focused on the analysis of survey answers. As the need for experimental data grew in order to adequately investigate MECs occurring in response to specific stimuli, the methods used in lab or online studies became increasingly diverse. These methods are described in this section, with a focus on self-reports and objective measures of MECs (summarised in Table 2.1), as well as experimental paradigms (summarised in Table 2.2) which have dominated the empirical literature on MECs.

Self-reports

When listening to music, MECs can either be self-reported or observed, and recorded retrospectively or continuously. A popular and convenient way to measure MECs is to rely completely on retrospective self-reports about the frequency or intensity of MECs (see Table 2.1 for a list of papers using this approach), generally collected with a short questionnaire after each trial. This has the advantage of requiring virtually no resources, but is also one of the least informative ways to record MECs. As a more detailed approach, continuous self-reports allow researchers to collect data on the specific timing of the onset—and sometimes offset—of MECs, with the exception of two studies in which participants were asked to keep a count of experiences of MECs on a scratch sheet (Balteș et al., 2011; Balteș & Miu, 2014). In their simplest form, continuous self-reports can be collected by asking participants to raise their finger or hand for the duration of experienced MECs (Craig, 2005; Goldstein, 1980; Konečni et al., 2007; Panksepp, 1995). Most commonly, however, participants report MECs by pressing on a button (see Table 2.1), sometimes in conjunction with continuous self-reports of valence and arousal, using bespoke interfaces such as *EMuJoy* (Nagel et al., 2007). In a few cases, an analogue slider (Bannister & Eerola, 2018) or a pressure-sensitive handle (Grunkina et al., 2017; Klepzig et al., 2020) have been used instead of a button to collect continuous ratings of MECs intensity, rather than a binary response about the occurrence of MECs.

An important methodological consideration in studies that use button presses for MECs and collect skin conductance response data is whether the act of pressing a button raises skin conductance response by itself. This has been consistently demonstrated not to be the case (Bannister, 2020c; Colver & El-Alayli, 2016; Grewe et al., 2011; Grewe et al., 2009a; Grewe et al., 2007; Guhn et al., 2007; Mori & Iwanaga, 2014b, 2015; Rickard, 2004; Salimpoor et al., 2009). Relatedly, several studies have validated button presses by only including the reported MECs in the analysis if they are accompanied by an increase in skin conductance response (Bannister, 2020c; Beier et al., 2020; Colver & El-Alayli, 2016; Egermann et al., 2011; Grewe et al., 2007; Mori & Iwanaga, 2014b). This approach has the advantage of not exclusively relying on self-reports, but considering the current lack of understanding regarding the exact relationship between MECs and skin conductance response (see Section 2.3.4), it might also lead to valid occurrences of MECs being discarded, depending on the chosen threshold.

Objective measures

The ideal way to record MECs would consist of an objective and continuous measure. Panksepp and Bernatzky (2002) made a brief reference to an inconclusive attempt at measuring MECs using thermal imaging of the skin surface, following a suggestion to use objective measures in an earlier publication (Panksepp, 1995). The authors concluded that directly measuring piloerection might be more appropriate, as previously suggested by Sloboda (1991). This can be done manually, as was the case in a study in which participants placed their arm through a curtain, and observers noted the onset and offset of piloerection (Craig, 2005), or automatically, using devices which can monitor piloerection.

The most notable example of such devices is the *Goosecam* (Benedek et al., 2010), an optical device which can be roughly described as a camera embedded in a box that blocks external light, recording the skin of the forearm—or lower leg in some later studies—from a close distance. LED lights shine on the skin at an angle from within the box, allowing goosebumps to cast a shadow on the skin. Images are then processed with a MATLAB toolbox using a discrete Fourier transform to provide a continuous measure of piloerection. A piloerection event occurs if the computed value exceeds an arbitrarily set threshold—usually defined in terms of the number of standard deviations away from a baseline recording—for a specified number of consecutive frames. The Goosecam has been tested in one participant who had voluntary control over piloerection (for an interesting exploratory investigation of this phenomenon, see Heathers et al., 2018), and was found to provide observations consistent with human judges (Benedek et al., 2010). It has since been used in several studies (see Table 2.1, as well as Chapter 3 for use of the Goosecam in the present research).

Another piloerection-monitoring device was proposed by Kim et al. (2014), and consists of a very thin, flexible, and compact sensor made of conductive polymer, which can be affixed to the skin to measure the physical deformation of its surface when goosebumps occur. The device was tested and validated by the authors, but while it represents an elegant solution, it remains unused in other studies to date, possibly because it requires resources which are less accessible than those needed to build a Goosecam.

Table 2.1: Measures of MECs

Type	Method	Papers
Retrospective self-reports		Bannister (2019), Blood and Zatorre (2001), Carr and Rickard (2016), Chabin et al. (2020), Goodchild et al. (2019), Honda et al. (2020), Jaimovich et al. (2013), Ji et al. (2019), Juslin et al. (2014), Park et al. (2019), Polo (2017), Schäfer and Sedlmeier (2011), Schoeller and Perlovsky (2016), Schoeller, Eskinazi, and Garreau (2018), Schoeller and Eskinazi (2019), Seibt et al. (2017), Silvia et al. (2015), Solberg and Dibben (2019), Strick et al. (2015), Wassiliwizky et al. (2015), Weth et al. (2015)
Continuous self-reports	Raising finger or hand	Craig (2005), Goldstein (1980), Konečni et al. (2007), Panksepp (1995)
	Scratch sheet	Balteș et al. (2011), Balteș and Miu (2014)
	Button	Bannister (2020c), Beier et al. (2020), Colver and El-Alayli (2016), Egermann et al. (2011), Ferreri et al. (2019), Grewe et al. (2007), Grewe et al. (2009a), Grewe et al. (2011), Guhn et al. (2007), Laeng et al. (2016), Mas-Herrero et al. (2014), Mori and Iwanaga (2014b), Mori and Iwanaga (2015), Mori and Iwanaga (2017), Nagel et al. (2008), Polo (2017), Rickard (2004), Sachs et al. (2016), Salimpoor et al. (2009), Salimpoor et al. (2011), T. W. Schubert et al. (2018), Seibt et al. (2018), Starcke et al. (2019), Sutherland et al. (2009), Wassiliwizky, Koelsch, et al. (2017), Zickfeld, Schubert, Seibt, Blomster, et al. (2019)
	Analogue slider	Bannister and Eerola (2018)
	Pressure-sensitive handle	Grunkina et al. (2017), Klepzig et al. (2020)
Objective measures	Thermal imaging (inconclusive)	Panksepp and Bernatzky (2002)
	Direct observation	Craig (2005)
	Goosecam	Benedek et al. (2010), Benedek and Kaernbach (2011), Quesnel and Riecke (2018), Sumpf et al. (2015), Wassiliwizky, Jacobsen, et al. (2017), Wassiliwizky, Koelsch, et al. (2017)
	Conductive polymer sensor	Kim et al. (2014)

Paradigms

Careful study design is required to investigate the different aspects of MECs. A popular approach initially used by Blood and Zatorre (2001) and in many later studies (see Table 2.2) requires participants to provide songs during which they often experience MECs. They are then asked to listen to these songs and to songs provided by other participants, which act as a control. This has the clear advantages of ensuring that genuine MECs are experienced, and excluding the possibility that the effects observed were simply due to the properties of each piece of music, since one participant's MECs-inducing stimulus is another participant's control stimulus. Common findings in these studies are that participants experience more MECs when listening to self-selected music, highlighting possible effects of familiarity, stylistic preference, and meaning (see Sections 2.3.6 and 2.3.7), and demonstrating that MECs are not caused by stimulus-driven properties alone. While this study design has been particularly fruitful because MECs are often considered to be highly idiosyncratic (Nusbaum et al., 2014;

Panksepp, 1995), it is important to bear in mind that MECs most likely involve an interaction between listener, context, and music (see Section 2.4).

Other studies have compared or combined responses to self-selected stimuli and to stimuli selected by the researchers (either arbitrarily or following a pre-selection procedure), used experimenter-selected stimuli only, or participant-selected stimuli only (see Table 2.2). Each of these approaches have distinct advantages and disadvantages, such as the degree of control over what the participants listen to, or how familiar they are with each piece of music. More specifically, experimenter-selected stimuli allow precise control over stimulus properties and familiarity, but may not always elicit MECs, whereas participant-selected stimuli are very likely to induce genuine MECs, at the cost of lower control over stimulus properties or familiarity.

Other paradigms provide better opportunities for making precise causal inferences, through direct manipulation of the stimuli (Bannister, 2020c; Bannister & Eerola, 2018; Honda et al., 2020; Juslin et al., 2014; Park et al., 2019), administration of substances thought to alter the experience of MECs (Ferreri et al., 2019; Goldstein, 1980; Starcke et al., 2019), repeated presentation of the same stimuli to the same participant (Grewe et al., 2007), or more broadly, through the a priori design of clearly distinct experimental conditions (see Table 2.2). Note that here, we are referring to causal paradigms, and not necessarily to knowledge about what causes MECs, which is why these studies are discussed in different sections of this chapter based on how relevant their findings are to each section. Such causal designs are clearly capable of providing more robust insight into MECs than experiments providing only correlational evidence, although they come with their own set of challenges, such as manipulating stimuli while maintaining ecological validity and avoiding the introduction of confounding factors.

While less relevant to this review, it is worth mentioning a small set of studies that have used MECs as an independent variable, leading to findings that MECs led to improved communication and heightened self-perception in a music therapy context (Lee, 2008), as also hypothesised by Tihanyi (2016), had no effect on memory performance as measured by image recall (Carr & Rickard, 2016) or on craving reduction in abstinent individuals with alcohol use disorder (Mathis & Han, 2017), had an effect on gait, as seen by increased cadence and stride length, and reduced stride time (Park et al., 2019), did not improve mood or increase generosity, helpfulness, or prosocial behaviour (Konečni et al., 2007), but contradictorily, did promote altruistic behaviour (Fukui & Toyoshima, 2014). Three devices have also been designed in an attempt to induce chills, through electrostatic force (Fukushima & Kajimoto, 2012) or coldness (Ishikawa et al., 2019; Schoeller et al., 2019), with the purpose of enhancing emotional

Table 2.2: Experimental paradigms used in research on MECs

Type	Design	Papers
No manipulation	Experimenter-selected music only	Balteș et al. (2011), Balteș and Miu (2014), Bannister (2019), Colver and El-Alayli (2016), Grewe et al. (2011), Grunkina et al. (2017), Guhn et al. (2007), Jaimovich et al. (2013), Ji et al. (2019), Klepzig et al. (2020), Konečni et al. (2007), Polo (2017), Schäfer and Sedlmeier (2011), T. W. Schubert et al. (2018), Seibt et al. (2017), Seibt et al. (2018), Silvia et al. (2015), Solberg and Dibben (2019), Strick et al. (2015), Wassiliwizky et al. (2015), Zickfeld, Schubert, Seibt, Blomster, et al. (2019)
	Participant-selected music only	Craig (2009), Fukui and Toyoshima (2013), Wassiliwizky, Jacobsen, et al. (2017)
	Participant- vs. experimenter-selected music	Benedek and Kaernbach (2011), Carr and Rickard (2016), Craig (2005), Grewe et al. (2007), Mas-Herrero et al. (2014), Nagel et al. (2008), Panksepp (1995), Quesnel and Riecke (2018), Rickard (2004), Weth et al. (2015), Wassiliwizky, Koelsch, et al. (2017)
	Participant-selected vs. other participants' music	Blood and Zatorre (2001), Laeng et al. (2016), Mori and Iwanaga (2014b), Mori and Iwanaga (2015), Mori and Iwanaga (2017), Sachs et al. (2016), Salimpoor et al. (2009), Salimpoor et al. (2011), Sumpf et al. (2015)
Manipulation	Stimulus manipulation	Bannister and Eerola (2018), Bannister (2020c), Honda et al. (2020), Juslin et al. (2014), Park et al. (2019)
	Stimulus comparison	Beier et al. (2020), Goodchild et al. (2019)
	Group comparison	Beier et al. (2020), Grewe et al. (2009a)
	Treatment comparison	Egermann et al. (2011), Schoeller and Perlovsky (2016), Schoeller, Eskinazi, and Garreau (2018), Sutherland et al. (2009)
	Longitudinal	Grewe et al. (2007)
	Neurochemical	Ferreri et al. (2019), Goldstein (1980), Starcke et al. (2019)
Other	Chills as independent variable	Carr and Rickard (2016), Fukui and Toyoshima (2014), Konečni et al. (2007), Lee (2008), Mathis and Han (2017), Park et al. (2019)
	Chills induction through physical means	Fukushima and Kajimoto (2012), Ishikawa et al. (2019), Schoeller et al. (2019)

experiences.

2.3.4 Physiological correlates

Being involved in emotional reactions, MECs are associated with autonomic nervous system activity (Kreibig, 2010), and are therefore accompanied by a set of physiological responses which have been studied extensively. We review these responses by examining how electrodermal, cardiac, and other physiological measures are associated with MECs (see Table 2.3 for a summary).

Skin measures

Electrodermal activity is typically decomposed into its tonic component, skin conductance level, reflecting slow, smooth changes in baseline activity, and its phasic component, skin conductance response, reflecting rapidly changing, event-related activity. Skin conductance level was found to increase around the onset of MECs, either shortly before they occur (Grewe et al., 2009a) or shortly after (Benedek & Kaernbach, 2011; Mori & Iwanaga, 2017), though a comparable number of studies found no effects of MECs on this measure (Balteș et al., 2011; Carr & Rickard, 2016; Jaimovich et al., 2013; Schäfer & Sedlmeier, 2011). The consensus is much more pronounced for skin conductance response, with many studies reporting associations with MECs (see Table 2.3), and only three not detecting such associations (Blood & Zatorre, 2001; Carr & Rickard, 2016; Jaimovich et al., 2013). Specifically, skin conductance response has been found to increase shortly before (Egermann et al., 2011; Grewe et al., 2009a; Salimpoor et al., 2009) or after (Benedek & Kaernbach, 2011; Grewe et al., 2011; Mori & Iwanaga, 2017) the onset of MECs, and to peak during (Craig, 2005; Salimpoor et al., 2009) or shortly after (Grewe et al., 2009a; Mori & Iwanaga, 2017) MECs. In some of these studies, however, self-reported MECs were only considered for analysis if accompanied by an increase in skin conductance response (see Section 2.3.3), which might have biased the results to some extent. Finally, peripheral skin temperature was found in some studies to decrease during MECs (Salimpoor et al., 2009) or with MECs intensity (Salimpoor et al., 2011), although others found no such association (Blood & Zatorre, 2001; Craig, 2005; Rickard, 2004).

Heart measures

Increases in heart rate (or decreases in interbeat interval—an inversely related variable) have generally been found to be associated with MECs, though, again, these findings have not always been replicated (see Table 2.3). Interestingly, in one study, heart rate was found to increase only for MECs that involve piloerection (Sumpf et al., 2015). Decreases in blood volume pulse amplitude (Benedek & Kaernbach, 2011; Salimpoor et al., 2011; Salimpoor et al., 2009), increases in E_K, a specific ratio of cardiac amplitudes in the resting electrocardiogram associated with emotionality (Sumpf et al., 2015), respiratory sinus arrhythmia, and power in the low frequency of heart rate variability (Balteș et al., 2011) have also been associated with MECs, while no effects were found for heart rate variability (Carr & Rickard, 2016), systolic blood pressure, diastolic blood pressure, power in the very low frequency of heart rate variability, and the ratio between low and high frequency powers of heart rate variability (Balteș et al., 2011).

Other measures

Empirical evidence is mixed on the relationship between MECs and an increase in respiration rate, with some studies finding supporting evidence, and others failing to identify such a relationship (see Table 2.3). Respiration depth, however, has been found to increase in all (Benedek & Kaernbach, 2011; Blood & Zatorre, 2001; Grewe et al., 2009a) but one study (Mori & Iwanaga, 2017). Muscle tension, as measured by electromyography, increased when listening to self-selected music known to induce MECs (Blood & Zatorre, 2001), but was not reported to increase with increased frequency of MECs (Rickard, 2004). Salivary cortisol levels decreased when listening to music that induces MECs (Fukui & Toyoshima, 2013) but not with increased frequency of MECs (Rickard, 2004). Other salivary hormone levels showed different patterns, with increases in estradiol, and no changes in testosterone, though it is important to note that this was in response to listening to music self-selected as likely to elicit MECs—occurrence of MECs was not actually recorded in this study (Fukui & Toyoshima, 2013). Pupil diameter, a physiological response associated with autonomic nervous system changes, increased during MECs, but this was not the case for eye blinks, saccade amplitude, or saccade dispersion (Laeng et al., 2016), and resting physiological state, recorded as a pre-experiment baseline, was found to be associated with the number of MECs when listening to self-selected music (Mori & Iwanaga, 2014b).

2.3.5 Neural correlates

The neural correlates of MECs are discussed in many papers, particularly when referring to the results of Blood and Zatorre (2001) and Salimpoor et al. (2011). Some very thorough reviews explore the neuroscience of music and emotion in depth, with significant coverage of the neuroscientific literature on MECs (e.g., Archie et al., 2013; Brattico et al., 2013; Brattico & Pearce, 2013; Chanda & Levitin, 2013; Habibi & Damasio, 2014; Koelsch, 2010, 2014; Salimpoor & Zatorre, 2013; Schaefer, 2017; Zatorre, 2003, 2015; Zatorre & Salimpoor, 2013). Therefore, this section of the chapter presents a brief summary of the main findings, examining how MECs are associated with the basal ganglia and other neural structures, as well as results from lesion and neurochemical studies, and research on anhedonia (see Table 2.4 for a summary by structure).

Basal ganglia

Structures belonging to the basal ganglia—a group of subcortical nuclei associated with motor control, executive functions, habit formation, reward, and emotion, among other functions—have been repeatedly linked with MECs. In the dorsal striatum, increases in activation have been found in the putamen and left caudate nucleus when comparing music listening with and without the experience of pleasant MECs (Klepzig et al., 2020). Furthermore, in an earlier study, the right caudate nucleus showed increased activation in anticipation of MECs, as well as a positive relationship between dopamine release and number of

Table 2.3: Physiological correlates of MECs

System	Measure	Papers and findings
Skin	Skin conductance level	**Increase:** Benedek and Kaernbach (2011), Grewe et al. (2009a), Mori and Iwanaga (2017) **No effect:** Balteș et al. (2011), Carr and Rickard (2016), Jaimovich et al. (2013), Schäfer and Sedlmeier (2011)
	Skin conductance response	**Increase:** Bannister and Eerola (2018), Benedek and Kaernbach (2011), Craig (2005), Egermann et al. (2011), Grewe et al. (2007), Grewe et al. (2009a), Grewe et al. (2011), Guhn et al. (2007), Klepzig et al. (2020), Mas-Herrero et al. (2014), Mori and Iwanaga (2014b), Mori and Iwanaga (2015), Mori and Iwanaga (2017), Polo (2017), Rickard (2004), Sachs et al. (2016), Salimpoor et al. (2009), Salimpoor et al. (2011) **Effect (direction not specified):** Grewe et al. (2007) **No effect:** Blood and Zatorre (2001), Carr and Rickard (2016), Jaimovich et al. (2013)
	Peripheral skin temperature	**Decrease:** Salimpoor et al. (2009), Salimpoor et al. (2011) **No effect:** Blood and Zatorre (2001), Craig (2005), Rickard (2004)
Heart	Heart rate	**Increase:** Benedek and Kaernbach (2011), Blood and Zatorre (2001), Grewe et al. (2009a), Guhn et al. (2007), Mas-Herrero et al. (2014), Polo (2017), Sachs et al. (2016), Salimpoor et al. (2009), Salimpoor et al. (2011), Sumpf et al. (2015) **No effect:** Balteș et al. (2011), Carr and Rickard (2016), Grewe et al. (2011), Jaimovich et al. (2013), Mori and Iwanaga (2017), Rickard (2004), Schäfer and Sedlmeier (2011)
	Blood volume pulse amplitude	**Decrease:** Benedek and Kaernbach (2011), Salimpoor et al. (2009), Salimpoor et al. (2011)
	Lesser-used measures	See Section 2.3.4
Other	Respiration rate	**Increase:** Balteș et al. (2011), Salimpoor et al. (2009), Salimpoor et al. (2011) **No effect:** Benedek and Kaernbach (2011), Grewe et al. (2011), Mori and Iwanaga (2017), Sumpf et al. (2015)
	Respiration depth	**Increase:** Benedek and Kaernbach (2011), Blood and Zatorre (2001), Grewe et al. (2009a) **No effect:** Mori and Iwanaga (2017)
	Muscle tension	**Increase:** Blood and Zatorre (2001) **No effect:** Rickard (2004)
	Salivary cortisol	**Decrease:** Fukui and Toyoshima (2013) **No effect:** Rickard (2004)
	Lesser-used measures	See Section 2.3.4

MECs (Salimpoor et al., 2011). Effects have also been found in the ventral striatum, which showed increased activation in response to pleasant MECs in a healthy control, but not in a patient with lesions following an extended stroke of the left middle cerebral artery (Grunkina et al., 2017). Activation in the left ventral striatum increased when listening to music that was self-selected to elicit pleasant emotional responses, including MECs, and was positively correlated with ratings of MECs intensity (Blood & Zatorre, 2001). Within the ventral striatum, the right nucleus accumbens showed increased activation during MECs, and a positive relationship between dopamine release, intensity of MECs, and degree of pleasure (Salimpoor et al., 2011), suggesting an involvement of this structure in processing the hedonic and reinforcing aspects of musical pleasure (Chanda & Levitin, 2013).

Other subcortical structures and cortical regions

In addition to the nucleus accumbens, associations with MECs have been reported for a wide range of limbic and paralimbic structures (i.e., structures originating from brain areas typically associated with emotion, long-term memory, and motivation, among other functions), such as the amygdala (Griffiths et al., 2004; Grunkina et al., 2017) and the left hippocampus, both of which showed decreased activation as MECs intensity increased (Blood & Zatorre, 2001), as well as the cingulate cortex (Blood & Zatorre, 2001), the insular cortex (Blood & Zatorre, 2001; Griffiths et al., 2004; Grunkina et al., 2017; Klepzig et al., 2020), and the orbitofrontal cortex (Blood & Zatorre, 2001), which all displayed increased activation with MECs (or an impaired ability to experience MECs for patients with an insular lesion—see next subsection), demonstrating a widespread involvement of the limbic system and associated cortical regions. Other brain structures and cortical regions have also shown increased activation with MECs, such as the primary auditory cortex and the secondary somatosensory cortex (Grunkina et al., 2017), the thalamus (Blood & Zatorre, 2001; Grunkina et al., 2017; Klepzig et al., 2020), the dorsomedial midbrain, the supplementary motor area, the cerebellum (Blood & Zatorre, 2001), including the right cerebellar hemisphere (Klepzig et al., 2020), and the locus coeruleus, as indicated by pupillary dilation during MECs (Laeng et al., 2016), as well as decreased activation for the ventromedial prefrontal cortex, the cuneus, and the precuneus (Blood & Zatorre, 2001).

Structural, neuropsychological, and neurochemical findings

White matter connectivity, the volume or density of the myelinated pathways between different areas of the brain, was investigated by Sachs et al. (2016), who reported increased tract volume from the posterior superior temporal gyrus to the anterior insula and medial prefrontal cortex—these tracts being part of the uncinate fasciculus, among others—in people who experience MECs frequently and consistently, but no difference in corticospinal tract volume, suggesting that these differences are specific, and not a result of general differences in white matter connectivity (Sachs et al., 2016). A study taking advantage of data from the Human Connectome Project (Van Essen et al., 2013) revealed that proneness to MECs is associated with higher resting-state functional connectivity between the default network and sensory and motor cortices, between the ventral default and salience networks, and lower connectivity between the cerebellum and somatomotor cortex, suggesting a greater integration between environmental perception and internal emotional experience (Williams et al., 2018).

Lesion studies have provided support for the involvement of these structures and tracts. A patient with lesions in the left insula and left amygdala exhibited impaired emotional processing of music, despite normal music perception and processing (Griffiths et al., 2004). Another patient lost the ability to perceive subtle differences between musical performances and to experience pleasure and MECs, following a lesion in the right putamen that impaired connectivity between the right insula and the superior temporal lobe, including the auditory cortex (Satoh et al., 2016). Finally, another patient with damage in the pyramidal tract, uncinate fasciculus, and left anterior insular cortex showed reports of MECs intensity consistent with a healthy control, but diminished bodily responses as indexed by changes in skin conductance level and skin conductance response (Grunkina et al., 2017).

Neurochemical findings provide some clarity on the role of endogenous opioids and dopamine. MECs were attenuated in three out of ten participants administered with naloxone, an opiate receptor antagonist (Goldstein, 1980)—a preliminary finding which received further support from a decrease in self-reported pleasure for pleasurable music after inducing anhedonia with naltrexone, a μ-opioid antagonist similar to naloxone (Mallik et al., 2017). Furthermore, the amount of time experiencing MECs was higher than placebo following intake of levodopa, a dopamine precursor, and lower than placebo following intake of risperidone, a dopamine antagonist (Ferreri et al., 2019).

Anhedonia

The literature on anhedonia further supports the results of neuroimaging, neurochemical, and lesion studies. Higher physical anhedonia, characterised by diminished reward from physical and sensory experiences, has been associated with experiencing MECs less often (Nusbaum et al., 2015), and shown to involve reduced activation in the left ventral striatum and increased activation in the ventromedial cortex (Dowd & Barch, 2012; Harvey et al., 2007; as cited by Nusbaum et al., 2015). Specific musical anhedonia, characterised by a failure to find music rewarding despite normal music perception, normal musical emotion recognition, and the absence of generalised anhedonia, can be measured with the Barcelona Musical Reward Questionnaire (Mas-Herrero et al., 2013), and has been found to be associated with fewer and less intense experiences of MECs, and a lack of increase in skin conductance response (except for one anhedonic participant), despite behavioural reports of MECs by some anhedonic participants (Mas-Herrero et al., 2014). Interestingly, tract volume between the left superior temporal gyrus and the left nucleus accumbens was shown to be lower for participants with severe musical anhedonia (Loui et al., 2017), providing further support for the involvement of white matter connectivity between auditory and

Table 2.4: Neural correlates of MECs

Group	Structure	Papers and findings
Basal ganglia	Dorsal striatum: Putamen	**Increased activation:** Klepzig et al. (2020) **Impaired with right lesion:** Satoh et al. (2016)
	Dorsal striatum: Caudate nucleus	**Increased left activation:** Klepzig et al. (2020) **Increased right activation:** Salimpoor et al. (2011)
	Ventral striatum	**Increased activation:** Grunkina et al. (2017) **Increased left activation:** Blood and Zatorre (2001)
	Ventral striatum: Nucleus accumbens	**Increased right activation:** Salimpoor et al. (2011)
Limbic and paralimbic structures	Amygdala	**Decreased activation:** Blood and Zatorre (2001) **Increased activation:** Grunkina et al. (2017) **Impaired with left lesion:** Griffiths et al. (2004)
	Hippocampus	**Decreased left activation:** Blood and Zatorre (2001)
	Cingulate cortex	**Increased activation:** Blood and Zatorre (2001)
	Insular cortex	**Increased activation:** Blood and Zatorre (2001), Grunkina et al. (2017), Klepzig et al. (2020) **Impaired with left lesion:** Griffiths et al. (2004), Grunkina et al. (2017)
	Orbitofrontal cortex	**Increased activation:** Blood and Zatorre (2001)
Other	Primary auditory cortex	**Increased activation:** Grunkina et al. (2017)
	Secondary somatosensory cortex	**Increased activation:** Grunkina et al. (2017)
	Ventromedial prefrontal cortex	**Decreased activation:** Blood and Zatorre (2001)
	Thalamus	**Increased activation:** Blood and Zatorre (2001), Grunkina et al. (2017), Klepzig et al. (2020)
	Dorsomedial midbrain	**Increased activation:** Blood and Zatorre (2001)
	Supplementary motor area	**Increased activation:** Blood and Zatorre (2001)
	Cerebellum	**Increased activation:** Blood and Zatorre (2001) **Increased right activation:** Klepzig et al. (2020)
	Locus coeruleus	**Increased activation:** Laeng et al. (2016)
	Cuneus	**Decreased activation:** Blood and Zatorre (2001)
	Precuneus	**Decreased activation:** Blood and Zatorre (2001)
Tracts	Uncinate fasciculus	**Impaired with lesion:** Grunkina et al. (2017) **Increased tract volume:** Sachs et al. (2016)
	Pyramidal tract	**Impaired with lesion:** Grunkina et al. (2017) **No effect of corticospinal tract volume:** Sachs et al. (2016)
	Right insula to superior temporal lobe	**Impaired with lesion:** Satoh et al. (2016)
	Default network to sensory and motor cortices	**High functional connectivity:** Williams et al. (2018)
	Ventral default to salience network	**High functional connectivity:** Williams et al. (2018)
	Cerebellum to somatomotor cortex	**Low functional connectivity:** Williams et al. (2018)

limbic structures.

2.3.6 Elicitors

The stimulus-driven elicitors of MECs fall into three broad categories: low-level acoustic elicitors, representing basic properties of the auditory signal, high-level musical elicitors, representing stimulus properties more specific to music, such as harmonic movement, and emotional elicitors, representing subjectively felt emotions in pieces of music. Understanding these elicitors is necessary in order to assess which psychological mechanisms might underlie MECs, and to inform theories on the function of MECs. As a result, considerable attention has been given to identifying these elicitors, as reviewed in this section, and summarised in Table 2.5.

Acoustic elicitors

MECs have repeatedly been linked with dynamic acoustic changes, and most often when such changes are sudden (Auricchio, 2017; Guhn et al., 2007; Nagel et al., 2008; Polo, 2017; Sloboda, 1991). More specifically, increased loudness or more frequent peaks in loudness were found around the onset of MECs (Beier et al., 2020; Grewe et al., 2007; Guhn et al., 2007; Honda et al., 2020; Nagel et al., 2008), particularly in the 920–4400 Hz band (Nagel et al., 2008). Loudness was also associated with continuous ratings of MECs intensity (Bannister & Eerola, 2018), and experimentally increasing the loudness of a musical passage known to often induce MECs and likely to engage *auditory looming* (see Section 2.3.8) resulted in more frequent experiences of MECs (Bannister, 2020c). Pleasure could be a mediating factor, however, with changes in volume leading to increased pleasure in some cases (Grewe et al., 2007), but decreased in others (Bannister, 2020c). MECs have also been shown to co-occur with higher event density (Bannister & Eerola, 2018; Nagel et al., 2008; Polo, 2017), expansion of the frequency range in the high or low register (Guhn et al., 2007; Polo, 2017), higher spectral centroid and spectral flux (Bannister & Eerola, 2018), increased roughness, dissonance, or fluctuation strength (Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Nagel et al., 2008; Park et al., 2019), higher variance in interaural level difference, a measure which captures rotation in binaural recordings (Honda et al., 2020), and increased sharpness or brightness (Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Honda et al., 2020), although, for one specific song, increasing brightness was found to reduce the frequency of MECs (Bannister, 2020c).

Musical elicitors

A number of features more specific to music have also been identified as potential elicitors of MECs, expanding on what was initially described as “dramatic peaks and valleys in music” (Goldstein, 1980, p. 127). Related to increases in loudness discussed in the previous paragraph, crescendi, build-ups, and climaxes have been linked with MECs (Auricchio, 2017; Bannister, 2020b; Bannister & Eerola, 2018; Panksepp, 1995; Polo, 2017; Solberg & Dibben, 2019). In addition to sudden dynamic changes, Sloboda (1991) identified several structural characteristics of musical excerpts that elicit MECs, such as new or unprepared harmonies, sudden textural changes, melodic appoggiaturas, enharmonic changes, specific melodic or harmonic sequences, or prominent musical events arriving earlier than prepared for, among others. Similar melodic and harmonic properties, including structural transitions and alterations such as changes in tonality, were subsequently associated with MECs in several empirical studies (Auricchio, 2017; Bannister, 2020b; Bannister & Eerola, 2018; Guhn et al., 2007; Mlejnek, 2013; Schurtz et al., 2012), in addition to rhythmic properties (Schurtz et al., 2012; Solberg & Dibben, 2019), although the two latter studies lack specific detail about which rhythmic properties were involved (for a hypothesis about optimal tempo, see McEvilly, 1999). A recurrent theme is textural changes (Auricchio, 2017; Polo, 2017; Sloboda, 1991; Solberg & Dibben, 2019), particularly with the entrance of new instruments, and the alternation, contrast, or communion between solo and accompanying instruments (Auricchio, 2017; Bannister, 2020b; Bannister & Eerola, 2018; Goodchild et al., 2019; Guhn et al., 2007; Mlejnek, 2013), which are considered particularly pleasurable by listeners (Grewe et al., 2007). Voice and lyrics have also been identified as potent elicitors of MECs (Bannister, 2020b; Schurtz et al., 2012), and some researchers have identified passages from slow movements (Guhn et al., 2007) and virtuosity (Mlejnek, 2013) as possible causes of MECs.

Finally, in a causal study by Bannister and Eerola (2018), MECs were found to happen less frequently, and to be rated as less intense, when specific passages known to often elicit MECs were removed from three pieces of music. Interestingly, as opposed to MECs, skin conductance response did not diminish when these passages were removed. This suggests that physiological arousal is dependent on local musical context, and possibly linked to the anticipation of MECs. Another point of interest reported by Bannister and Eerola (2018) is that acoustic and musical elicitors might be intrinsically related, since the entrance of new instruments, for instance, would naturally come along with dynamic and spectral changes (see also Auricchio, 2017). Research that comprehensively teases apart the effects of acoustic and musical elicitors is needed to better understand how stimulus properties influence the occurrence of MECs.

Emotional elicitors

MECs can also arise from the perception of emotions expressed by music, which, for present purposes, can be broadly grouped into valence, emotionality, and meaning. While frequency of self-reported MECs has been found to increase when listening to music rated as positively valenced (Grewe et al., 2011), associations between MECs and perceived sadness in female participants were found by Panksepp (1995) following a series of experiments. In this study, however, both happy and sad music were reported to elicit MECs, as was the case in other studies linking both positive and negative perceived emotions with MECs (Bannister, 2020b; Mori & Iwanaga, 2017). Rather than valence, greater perceived emotionality, whether positively or negatively valenced, has often been identified as a possible cause of MECs, whether it is referred to as such (Beier et al., 2020), as emotional power (Rickard, 2004), as perceived emotional content (Panksepp, 1995), as emotional intensity (Bannister & Eerola, 2018), or as the climactic stage of an aesthetic experience (Panzarella, 1980).

Finally, related to the effect of lyrics discussed in the previous subsection (Bannister, 2020b; Schurtz et al., 2012), MECs have been found to be associated with the perception of meaning in music, whether it is meaning of lyrics (Bannister, 2020b), personal meaning (Craig, 2009; Goldstein, 1980), or extra-musical meaning, such as pride or patriotism (Mlejnek, 2013). Notably, some studies of the effects of meaning have focused on priming effects, and resulted in conflicting perspectives. Specifically, while there was little to no effect of presenting various types of priming stimulus (national anthems, stories, architectural objects, paintings) on the frequency or duration of MECs when subsequently listening to a piece of music (Konečni et al., 2007), being exposed to a complex, existential statement, as opposed to an incoherent statement, increased the number of chills experienced when watching subsequent video clips (Schoeller, Eskinazi, & Garreau, 2018; Schoeller & Perlovsky, 2016). Interestingly, Konečni et al. (2007) also observed that there was no priming effect of experiencing MECs themselves on subsequent experiences of MECs, whereas frequency of MECs has been found to increase (Benedek & Kaernbach, 2011) or decrease (Laeng et al., 2016) with trial number during experiments (and therefore, following previous occurrences of MECs), highlighting a lack of consensus on the matter.

Underlying mechanisms

When it comes to understanding how these various elicitors might cause MECs, it is useful to consider potential underlying psychological mechanisms. A useful framework for doing so comes from an extensive body of work which sought to provide a unified theory of evoked musical emotions in the form of a set of underlying mechanisms (Juslin, 2013; Juslin & Västfjäll, 2008), the diversity of which was echoed by Huron (2016) when discussing the range of ways in which sounds are thought to evoke pleasure. It could be that these mechanisms are also involved in the experience of MECs, by evoking emotions which would in turn induce MECs, or by directly inducing MECs, but not fully-fledged emotional

Table 2.5: Elicitors of MECs

Category	Elicitor	Papers and findings
Acoustic	Loudness	**Sudden change:** Auricchio (2017), Guhn et al. (2007), Nagel et al. (2008), Polo (2017), Sloboda (1991) **Increase or more frequent peaks:** Bannister and Eerola (2018), Bannister (2020c), Beier et al. (2020), Grewe et al. (2007), Guhn et al. (2007), Honda et al. (2020), Nagel et al. (2008)
	Event density	**High levels:** Bannister and Eerola (2018), Nagel et al. (2008), Polo (2017)
	Frequency range	**Expansion in high or low register:** Guhn et al. (2007), Polo (2017)
	Spectral centroid or flux	**High levels:** Bannister and Eerola (2018)
	Roughness, dissonance, or fluctuation strength	**Increase:** Bannister and Eerola (2018), Beier et al. (2020), Grewe et al. (2007), Nagel et al. (2008), Park et al. (2019)
	Brightness or sharpness	**Increase:** Bannister and Eerola (2018), Beier et al. (2020), Grewe et al. (2007), Honda et al. (2020) **Decrease:** Bannister (2020c)
	Interaural level difference	**High variance:** Honda et al. (2020)
Musical	Crescendi, build-ups, and climaxes	Auricchio (2017), Bannister and Eerola (2018), Bannister (2020b), Goldstein (1980), Panksepp (1995), Polo (2017), Solberg and Dibben (2019)
	Changes in structure, melody, or harmony	Auricchio (2017), Bannister and Eerola (2018), Bannister (2020b), Guhn et al. (2007), Mlejnek (2013), Schurtz et al. (2012), Sloboda (1991)
	Rhythmic properties	Schurtz et al. (2012), Solberg and Dibben (2019)
	Textural changes	**In general:** Auricchio (2017), Polo (2017), Sloboda (1991), Solberg and Dibben (2019) **Entrance or interplay between instruments:** Auricchio (2017), Bannister and Eerola (2018), Bannister (2020b), Goodchild et al. (2019), Guhn et al. (2007), Mlejnek (2013)
	Voice and lyrics	Bannister (2020b), Schurtz et al. (2012)
	Slow movements	Guhn et al. (2007)
	Virtuosity	Mlejnek (2013)
Emotional	Perceived valence	**Positive:** Grewe et al. (2011) **Both positive and negative:** Bannister (2020b), Mori and Iwanaga (2017), Panksepp (1995)
	Perceived emotionality	Bannister and Eerola (2018), Beier et al. (2020), Grewe et al. (2009a), Panksepp (1995), Panzarella (1980), Rickard (2004)
	Perceived meaning	**Effect:** Bannister (2020b), Craig (2009), Goldstein (1980), Mlejnek (2013), Schoeller and Perlovsky (2016), Schoeller, Eskinazi, and Garreau (2018) **No effect:** Konečni et al. (2007)

experiences.

In this framework, *brain stem reflex* refers to the process by which low-level acoustic features quickly and automatically elicit emotions when exceeding a threshold value (Juslin, 2013), and would provide a reasonable explanation as to why acoustic elicitors such as sudden changes in loudness or dissonance might cause physiological arousal and MECs (L. Harrison & Loui, 2014; Juslin et al., 2014), although it is worth pointing out that for MECs, the corresponding mechanism reflects relatively automatic reactions to sudden changes in the acoustic signal, rather than a psychological startle response specifically. *Musical expectation*, as discussed in Chapter 1, is based on the hypothesis that developing expectations follows a process of probabilistic learning of the statistical regularities in musical structure (Pearce, 2018; Saffran et al., 1999). Musical expectation has often been posited as a cause of MECs (L. Harrison & Loui, 2014; Huron, 2006; Huron & Margulis, 2010; Juslin, 2013; Juslin & Västfjäll, 2008; McDermott, 2012; Mencke et al., 2019; Pearce & Wiggins, 2012; Salimpoor et al., 2011; Sloboda, 1991), and indeed, the majority of the musical elicitors discussed in this section could engage such a mechanism. Interestingly, Levinson (2006) suggested that there might be two types of MECs, the first type induced timbrally or dynamically, and the second type induced melodically, harmonically, or rhythmically. This is consistent with the possible involvement of brain stem reflex, on the one hand, and musical expectation, on the other.

Other mechanisms underlying emotional responses to music have also been discussed in relationship to MECs, such as *episodic memory* (Goldstein, 1980), *evaluative conditioning*, or *emotional contagion* (L. Harrison & Loui, 2014), all of which have been linked speculatively by these authors to some of the emotional elicitors discussed in the previous subsection. Paradoxically, when underlying mechanisms were explicitly investigated, either systematically (Juslin et al., 2014) or through self-reports (Bannister, 2020b; Bannister & Eerola, 2018), emotional contagion was strongly linked to MECs, but brain stem reflex and musical expectation were not. These results, however, could reflect the distinct possibility that the experimental manipulations of the musical stimuli did not adequately target the mechanisms in question, that listeners do not have sufficient conscious access to the reasons why they experience MECs to be able to self-report them, or that such conscious access varies between mechanisms. Further investigation is therefore needed to obtain conclusive answers about the psychological mechanisms that underlie MECs.

Associated factors

There exist other factors that potentially contribute to the elicitation of MECs. While these have rarely been the primary topic of investigation, they are often reported, and provide useful context to the findings discussed in this section. Some authors covered listening situations, comparing occurrences of MECs when listening to music alone or with others. In most cases, no differences were found (Egermann et al., 2011; Nusbaum et al., 2014; Sutherland et al., 2009), although peaks in skin conductance response were higher during MECs when listening alone than when listening in a group (Egermann et al., 2011), and survey respondents reported most experiences of MECs to happen during solo listening (Bannister, 2020b). These findings might reflect an effect of attention (Beier et al., 2020; Nusbaum et al., 2014; see also Mori & Iwanaga, 2014a), possibly related to alcohol intake being found to reduce frequency of MECs (Starcke et al., 2019), which would provide further support for the suggested role of attention in aesthetic responses (see Section 2.3.2). Interestingly, theories of dynamic attending (Jones & Boltz, 1989; Large & Jones, 1999) suggest a relationship between attention and temporal expectation, through which attention is directed at points in time which are expected to be more salient. Such a relationship could provide a possible mechanism through which increased attention affects MECs via musical expectation, and could provide a partial explanation for the involvement of neural structures associated with predictive timing and rhythm perception in MECs (Grahn & Brett, 2007; Teki et al., 2011), such as the basal ganglia. Future research should aim to investigate these issues in order to establish the precise nature of the relationship between attention and MECs.

Another important effect is that of repetition and familiarity. Listening to the same piece of music several times within a single experimental session was not found to affect the frequency or intensity of MECs (Balteș et al., 2011; Bannister, 2020c; Blood & Zatorre, 2001), but doing so every day over a week led to reduced frequency of MECs (Grewe et al., 2007), possibly due to habituation, although this longitudinal effect was investigated in only one participant. Over longer time scales, MECs have been reported to be a reliable response, and even to grow with repeated listening (Sloboda, 1991). More generally, conflicting effects of familiarity have been identified, with some studies reporting more occurrences of MECs for familiar stimuli (Craig, 2005; Grewe et al., 2009a; Panksepp, 1995; Rickard, 2004; Weth et al., 2015), and other studies reporting no effects of stimulus familiarity (Bannister, 2019; Bannister & Eerola, 2018; Benedek & Kaernbach, 2011; Colver & El-Alayli, 2016; Guhn et al., 2007; Rickard, 2004; Wassiliwizky et al., 2015), although some of these studies featured stimuli which were either all very familiar (Benedek & Kaernbach, 2011), or very unfamiliar (Colver & El-Alayli, 2016; Guhn et al., 2007). Familiarity has been argued to be a strong driver of aesthetic experiences, in conjunction with surprise, complexity, and expectation (Greasley & Lamont, 2016; Salimpoor et al., 2015; Verhaeghen, 2018), and could contribute to the elicitation of MECs by increasing recognition of meaning in music or by promoting a conflict between *schematic* and *veridical* expectation (Bharucha, 1994; Huron, 2006; Miranda & Ullman, 2007; Salimpoor et al., 2015), allowing unconscious surprise, caused by schematically unexpected events, to continue to occur in very familiar music, which would be veridically highly expected. This remains speculative, until further empirical research provides greater clarity on the association between familiarity and MECs.

2.3.7 Individual differences

While most people seem to have the ability to experience MECs (see Section 2.3.1), not everyone can or does so equally often. As a result, there has been some interest in identifying how individual differences might affect the prevalence of MECs and the frequency of experiencing them. In this section, we review the evidence on the role played by gender, age, musical training, and personality differences in the experience of MECs (see summary in Table 2.6).

Gender, age, and musical training

Panksepp (1995) identified in a series of experiments that women find sad music more likely to cause MECs than men, and vice versa for happy music, among other findings showing, especially for women, a relationship between MECs and perceived sadness. Similarly, Benedek and Kaernbach (2011) detected an effect of gender, with more women experiencing piloerection than men when listening to music and film audio, although the study involved an uneven gender ratio. The vast majority of studies that analysed the effect of gender, however, have reported no influence on MECs (see Table 2.6). The effect of age on MECs is less clear. Correlations with age have been found (Williams et al., 2018), including for some (e.g., goose pimples) but not all (e.g., shivers down the spine) reactions to music related to MECs (Mlejnek, 2013), and age positively predicted a small amount of variance in the number of MECs experienced during an opera performance (Balteș & Miu, 2014), whereas no effect of age was identified by Grewe et al. (2009a), Mori and Iwanaga (2014b), Starcke et al. (2019), and Zickfeld, Schubert, Seibt, Blomster, et al. (2019). Regarding effects of musical training, Nusbaum and Silvia (2011) found that playing an instrument is a significant predictor of the frequency of experiences of MECs, while Beier et al. (2020) reported effects of Western music theory knowledge on MECs experienced when listening to Western, Indian, but not Chinese music. However, other empirical evidence does not support an effect of musical training or musical sophistication (Müllensiefen et al., 2014) on MECs (Bannister & Eerola, 2018; Grewe et al., 2009a; Guhn et al., 2007; Polo, 2017; Rickard, 2004). It is important to note that most of these findings were not hypothesis-driven and there is very little theoretical basis for hypothesising effects of gender, age and musical training on MECs. Considering this limitation, as well as the limited scope of some of the results (discussed above), it is reasonable to assume that, for the most part, MECs are experienced independently of gender, age, and musical training.

Personality correlates

By far the most documented personality correlate of the experience of MECs is openness to experience—a Big Five personality trait characteristic of individuals who are curious, innovative, imaginative, sensitive to the arts, and who experience a wide range of feelings and emotions (McCrae, 2007). The relationship between MECs and openness to experience has been identified in many studies (see Table 2.6), though it was ambiguous in some cases (Mori & Iwanaga, 2015; Sumpf et al., 2015), and not present in others (Mathis & Han, 2017; Rickard, 2004; Starcke et al., 2019). Importantly, the NEO Personality Inventory and the NEO Five-Factor Inventory (Costa & McCrae, 1992) used in the majority of these studies both include an item about experiencing chills, which counts towards openness to experience. This raises the concern that the empirical relationship between MECs and openness to experience might be driven by the contribution of this item towards the scale. However, this seems not to be the case, because the item about chills is highly correlated with the sum of the remaining items on the openness to experience scale, as shown by corrected item-total correlations for this trait. Moreover, this analysis revealed that out of all items, the one about chills is the most highly correlated with the rest of the scale, making it the best cross-cultural indicator of openness to experience (McCrae, 2007). In addition, this item was confirmed to be related to the number of MECs experienced in a lab environment (Colver & El-Alayli, 2016). The other Big Five traits have also been investigated, and found to predict some of the variance in the frequency of experiencing MECs when taken together (Nusbaum & Silvia, 2011; Silvia & Nusbaum, 2011), and individually in the case of extraversion, neuroticism, and agreeableness (see Table 2.6), though the relationship for the latter has been found to be both positive (Sumpf et al., 2015; Williams et al., 2018) and negative (Maruskin et al., 2012). Agreeable individuals were also found to be more likely to experience MECs with piloerection rather than without (Sumpf et al., 2015).

Aside from the Big Five traits, many personality factors have been investigated. Experiencing MECs was found to be associated with being more observing and judging (N. R. Harrison & Clark, 2016), less susceptible to anger (Laeng et al., 2016), more likely to follow the *music-empathising* cognitive style of music listening, which is linked with a greater focus on emotional content (Linnemann et al., 2018), and more likely to listen to music in order to reduce negative affect rather than to stimulate fun (Starcke et al., 2019). There are conflicting results about the effects of reward dependence and sensitivity (Bannister, 2020c; Grewe et al., 2007; Mori & Iwanaga, 2015, see Section 2.3.5 for the relationship between MECs and anhedonia), thrill and adventure seeking (Grewe et al., 2007; Mathis & Han, 2017), stylistic preference (Bannister & Eerola, 2018; Nusbaum & Silvia,

Table 2.6: Individual differences in susceptibility to MECs

Type	Characteristic	Papers and findings
Demographic	Gender	**Effect:** Benedek and Kaernbach (2011), Panksepp (1995) **No effect:** Bannister (2019), Grewe et al. (2007), Grewe et al. (2009a), Guhn et al. (2007), Goldstein (1980), N. R. Harrison and Clark (2016), Mlejnek (2013), Mori and Iwanaga (2014b), Polo (2017), Rickard (2004), Silvia and Nusbaum (2011), Starcke et al. (2019), Sutherland et al. (2009), Williams et al. (2018), Zickfeld, Schubert, Seibt, Blomster, et al. (2019)
	Age	**Effect:** Balteș and Miu (2014), Mlejnek (2013), Williams et al. (2018) **No effect:** Grewe et al. (2009a), Mori and Iwanaga (2014b), Starcke et al. (2019), Zickfeld, Schubert, Seibt, Blomster, et al. (2019)
Experiential	Musical training	**Effect:** Beier et al. (2020), Nusbaum and Silvia (2011) **No effect:** Bannister and Eerola (2018), Grewe et al. (2009a), Guhn et al. (2007), Polo (2017), Rickard (2004)
Personality	Big Five	**Effect:** Nusbaum and Silvia (2011), Silvia and Nusbaum (2011)
	Big Five: Openness	**Effect:** Bannister (2020b), Colver and El-Alayli (2016), Maruskin et al. (2012), McCrae (2007), Mori and Iwanaga (2015), Nusbaum and Silvia (2011), Silvia and Nusbaum (2011), Silvia et al. (2015), Sumpf et al. (2015) **Ambiguous:** Mori and Iwanaga (2015), Sumpf et al. (2015) **No effect:** Mathis and Han (2017), Rickard (2004), Starcke et al. (2019)
	Big Five: Extraversion	**Effect:** Maruskin et al. (2012), Rickard (2004), Sumpf et al. (2015), Williams et al. (2018)
	Big Five: Neuroticism	**Effect:** Maruskin et al. (2012), Silvia et al. (2015), Sumpf et al. (2015), Williams et al. (2018)
	Big Five: Agreeableness	**Positive effect:** Sumpf et al. (2015), Williams et al. (2018) **Negative effect:** Maruskin et al. (2012)
	Lesser-used characteristics	See Section 2.3.7

2011), and *aesthetic fluency*, a measure of expertise in the arts (N. R. Harrison & Clark, 2016; Silvia & Nusbaum, 2011), and no effects were detected for fluid intelligence (Silvia & Nusbaum, 2011), mood (Balteș & Miu, 2014), vividness of visual imagery (Balteș & Miu, 2014), or impulsive or anxious behaviour (Honda et al., 2020).

Interestingly, when distinguishing between different categories of chills, personality correlates differ. Goosetingles have been associated with extraversion, approach temperament, and positive emotionality, while coldshivers have been linked with neuroticism, avoidance temperament, and negative emotionality (Maruskin et al., 2012) Similarly, while there appears to be no effect of trait empathy on MECs as a single construct (Balteș & Miu, 2014; Bannister, 2020c), empathy has been found to be associated with moving chills, but not cold or warm chills (Bannister, 2019), echoing the results of a meta-analysis in which trait empathic concern, associated with the state of being moved, has been linked to chills (Zickfeld et al., 2017).

2.3.8 Theoretical perspectives on function

The evidence reviewed so far mostly addresses what MECs are, and how they are elicited, but there remains the broader question of why MECs occur. This final section surveys current theoretical perspectives about their origin. Most of these theories are expressed in terms of the evolutionary basis of chills and they tend to overlap partially to varying degrees while also generally possessing distinctive features. It is therefore important in the context of the present chapter to clearly and carefully delineate these theories on the function of MECs. All of the theories are speculative to some degree. It is because of their speculative nature that we are closing rather than opening this chapter with these. At this point in time, none of the theories reviewed below have sufficient experimental support to provide a robust platform for scaffolding and interpreting the empirical literature as a whole. However, having surveyed the existing empirical literature, there is value in considering the extent to which empirical results to date corroborate or refute the predictions of these theories and the experimental evidence required for more conclusive assessment. With these goals in mind, we evaluate in this section theories proposing that chills are associated with separation calls, the emotional state of being moved, peak arousal, contrastive valence, and knowledge instinct.

Separation call

The idea behind the *separation call* theory is that, in many animal species, separation calls are used to motivate parents to locate their offspring who might have become lost. According to the theory, this need for social reunion is driven by a feeling of coldness elicited by separation calls and leading to piloerection, potentially caused by an overlap between brain functions governing thermoregulation and social bonding, thereby providing an evolutionary explanation for the purpose of MECs (Panksepp, 1995, 2009; Panksepp & Bernatzky, 2002). This theory was proposed following early findings suggesting that MECs are more likely in women, with music that is familiar, perceived as sad, and includes high-pitched crescendi, which could be respectively accounted for by mothers being more susceptible to separation calls, by social attachment being a learned behaviour in mammals, by sadness due to potential loss providing the emotional context for potential reunion, and by the acoustic characteristics of separation calls, according to Panksepp (1995). As discussed earlier in this chapter, however, the effects of gender, familiarity, and stimulus valence are far from clear-cut, and the diversity in possible elicitors of MECs cannot be fully explained by a similarity with separation calls.

Some researchers have argued that chills are indeed related to closeness and social bonding (Bériachvili, 2016; Bicknell, 2007; Maruskin et al., 2012; Schoeller & Eskinazi, 2019), linked to physiological changes consistent with a state of sadness (Benedek & Kaernbach, 2011), and that there might be an overlap between thermoregulatory and social functions (for a brief review, see Bannister, 2019). However, critics of the separation call theory have argued that it fails to account for the possible existence of different types of chills (Levinson, 2006; see also Bannister, 2019; Maruskin et al., 2012) or for chills being experienced in response to a varied range of stimuli (Bériachvili, 2016; Sachs et al., 2018), that it is not consistent with the personality correlates of individuals most susceptible to experiencing MECs (McCrae, 2007), that there is a lack of clarity about which stimulus properties would reflect separation calls (Bannister, 2020b), and that there is no evidence for the occurrence of chills in response to separation calls in nonhuman primates (Altenmüller et al., 2013). Despite an attempt to provide causal support (briefly described in Panksepp & Bernatzky, 2002, pp. 143–144) for the separation call theory, it does not fully account for current findings in the literature, and a clearer consensus for its supporting evidence would be needed to consider this theory even as a partial, if not complete, explanation for the occurrence of MECs.

Being moved

Other theories on the function of MECs have proposed that they are related to the emotional state of *being moved*. Originating in an identified relationship between moving music and MECs (Goldstein, 1980; Panksepp, 1995; Panksepp & Bernatzky, 2002), the concept found itself included in the *aesthetic trinity* of Konečni (2005), which comprises awe, being moved, and chills. Within the framework of the aesthetic trinity theory, being moved is often accompanied by chills, although both responses can occur independently, and the rarer response, awe, is always accompanied by experiences of being moved and chills (Konečni, 2005, 2007, 2008, 2013; Konečni et al., 2007). There is empirical support for a relationship between awe and chills (Cotter et al., 2018; Ji et al., 2019; Maruskin et al., 2012; Quesnel & Riecke, 2018; Schurtz et al., 2012; Silvia et al., 2015), but despite claims that experiencing aesthetic awe results from an evolutionary process of sexual selection (Konečni, 2005), the theory fails to clearly outline mechanisms for the occurrence of MECs (Bériachvili, 2016; Branković, 2013).

In another line of research, being moved has been included in the construct named *kama muta*, which represents a positive feeling, often involving tears, chills, and a subjective feeling of warmth in the chest, as a result of experiencing or observing an increase in communal sharing or closeness, and is associated with trait empathic concern (Fiske et al., 2019; T. W. Schubert et al., 2018; Seibt et al., 2018; Zickfeld, Schubert, Seibt, Blomster, et al., 2019; Zickfeld et al., 2017, 2019). While the experience of kama muta is not restricted to music listening, the co-occurrence of MECs and tears, notably, is well documented in the music psychology literature (Bannister, 2019; Cotter et al., 2018; Mori & Iwanaga, 2017; Scherer et al., 2001; Strick et al., 2015).

More generally, there have been many theoretical (e.g., Menninghaus et al., 2015) and empirical (Bannister, 2019, 2020b; Bannister & Eerola, 2018; Benedek & Kaernbach, 2011; Eerola et al., 2016; Panksepp, 1995; Seibt et al., 2017; Strick et al., 2015; Vuoskoski & Eerola, 2017; Wassiliwizky, Jacobsen, et al., 2017; Wassiliwizky, Koelsch, et al., 2017; Wassiliwizky et al., 2015; Weth et al., 2015) associations between MECs and being moved, with additional links to liking and perceived sadness. The aesthetic trinity and kama muta frameworks do not propose fully-fledged mechanisms explaining the relationship between being moved and chills, and furthermore, there is little detail about the evolutionary mechanisms which could underlie that relationship. However, the extent of the discourse is such that it seemed appropriate to include the emotional state of being moved in this section, and fleshing it out in more detail should be considered as a promising avenue for future research.

Peak arousal

Motivated by a series of empirical findings (Grewe et al., 2009a; Grewe et al., 2007; Guhn et al., 2007; Rickard, 2004; as cited by Benedek & Kaernbach, 2011), the *peak arousal* hypothesis was proposed, advancing that MECs occur when a threshold in emotional and physiological arousal is exceeded (Benedek & Kaernbach, 2011). A closely related idea was first formulated by Blood and Zatorre (2001), who suggested that MECs can be experienced once a certain level of pleasure and emotional arousal is reached, and indeed, as discussed earlier, many empirical studies have subsequently used MECs as an indicator of pleasurable responses to music, and uncovered relationships between pleasure, subjective arousal, and MECs.

Similarly, MECs have been shown unequivocally to be associated with physiological arousal (see Section 2.3.4), but this theory posits more specifically that MECs are indicators of peak emotional and physiological arousal. While some studies have investigated the time-course of such peak responses, as discussed earlier, there is a lack of agreement about their specific timing with respect to the onset of MECs. Furthermore, little is known about whether or not peaks of arousal or pleasure can occur in the absence of MECs, which raises the question of whether MECs are a cause or a consequence of emotional and physiological arousal. In their study, Benedek and Kaernbach (2011) found some evidence consistent with the peak arousal hypothesis, but also suggested that rapid, shallow breathing during chills is required to further support the hypothesis. Such breathing patterns were not observed in their study, or in most studies of respiration rate and depth during experiences of MECs. Overall, the empirical data available to date do not clearly support or refute the peak arousal hypothesis, and further systematic study is needed in order to fully examine the time-course of emotional and physiological arousal, as well as pleasure, in the presence and absence of MECs.

Contrastive valence

It has also been proposed that MECs can be caused by musical expectations, most notably through a process called *contrastive valence*. This process relies on *ITPRA*, a theory of expectation proposed by Huron (2006), according to which responses to a situation are separated into *imagination* and *tension*, its pre-outcome components, and *prediction*, *reaction*, and *appraisal*, its post-outcome components. When listening to music, MECs are thought to occur when a rapid, unconscious fear response due to an unexpected outcome causes piloerection, which is subsequently followed by a neutral or positive conscious appraisal of musical sounds as a safe stimulus, leading to pleasure due to the positive contrast in valence between these two responses (Huron, 2006; Huron & Margulis, 2010). According to the theory, pleasurable chills in response to an unexpected outcome, musical or not, reflect an exaptation of vestigial thermoregulation and intimidation responses, drawing their adaptive value from promoting attention and information processing, rewarding and reinforcing learning when faced with inaccurate predictions, facilitating memory formation, and driving curiosity to detect new, surprising patterns, through the recruitment of the dopaminergic reward system, in order to promote more effective decision making, thereby leading to positive future outcomes (Altenmüller et al., 2013; Cantor, 2019; Grewe et al., 2007; Huron, 2006; Huron & Margulis, 2010; Maruskin et al., 2012; Wassiliwizky, Koelsch, et al., 2017). Although, following Huron (2006), we focus here on contrastive valence, we believe this causal explanation for MECs is potentially also compatible with other theories on the psychological mechanisms underlying the effects of expectation on emotion and pleasure, including theories invoking tension and resolution (Meyer, 1956), and learning progress (Gold et al., 2019).

As discussed previously, many empirical findings are consistent with a role of schematic and veridical expectation in the experience of MECs. There are also distinct subjective, physiological, and neural differences between pre-outcome and post-outcome reactions when experiencing MECs (Bannister & Eerola, 2018; Grewe et al., 2009a; Salimpoor et al., 2011; Wassiliwizky, Koelsch, et al., 2017), but these findings lack the temporal precision to fully support the exact time-course proposed by the ITPRA theory. In addition, while the relationship between expectation and pleasure has been explicitly investigated (Cheung et al., 2019; Gold et al., 2019), comparable studies have yet to be conducted on the relationship between expectation and MECs. Critics of the theory argue that the lack of a universal stimulus-response pattern for MECs renders fear unlikely to be the primary evolutionary cause of MECs (Bannister, 2020b; Grewe et al., 2007; Nagel et al., 2008). However, this fails to account for the fact that different individuals can experience fear in response to different stimuli, based on experience and circumstances. Moreover, if expectation is involved, we would expect to see individual differences due to stylistic enculturation (Pearce, 2018; and for partial support of an effect of stylistic knowledge, see Beier et al., 2020). As with the other theories reviewed so far, however, contrastive valence doesn't fully account for the experience of MECs, notably by failing to provide an explanation for MECs caused by the emotional *expressiveness* of music (Levinson, 2006) and emotional elicitors.

Related to a fear-based response due to expectation mechanisms, it has recently been proposed that *auditory looming* is a possible cause of MECs, presumably reflecting an adaptive need to perceive and signal an approaching threat (Bannister, 2019, 2020c; Bannister & Eerola, 2018). This theory, linked to the role of vigilance in expectation (Huron, 2006), could explain how crescendi and sudden increases in loudness might cause MECs, and has received recent support from an experiment showing that manipulating loudness affects the occurrence of MECs (Bannister, 2020c). However, the auditory looming theory does not naturally explain the pleasure often associated with MECs, and it remains to be determined whether or not this can be attributed to contrastive valence.

Knowledge instinct

According to the *knowledge instinct* theory (see Schoeller, Perlovsky, & Arseniev, 2018), humans are driven to learn by modifying mental representations in order to match patterns in perceived stimuli. *Knowledge acquisition* consists of the creation and improvement of these representations, and knowledge instinct is the fundamental motivation for knowledge acquisition. Emotions arise from satisfaction or dissatisfaction of knowledge instinct, or in other words, from the congruence or incongruence between bottom-up sensory signals and top-down mental models. Positive aesthetic emotions occur when congruence remains high, and when content at the top of the cognitive hierarchy is engaged, possibly resulting in chills and experiences of the sublime (Schoeller, Eskinazi, & Garreau, 2018; Schoeller & Perlovsky, 2016; Schoeller, Perlovsky, & Arseniev, 2018). In other words, chills can occur if stimuli that are relevant to important abstract concepts, such as meaning, are accurately predicted and understood. This theory has also been expressed in terms of an interaction between environment and encoded schema (Pelowski et al., 2017; Pelowski et al., 2018).

While this theory could account for the relationship between MECs and the perception of meaning, and has received tentative support from the effect of the coherence of a priming statement on subsequent experiences of chills when watching video clips (Schoeller, 2015; Schoeller, Eskinazi, & Garreau, 2018; Schoeller & Perlovsky, 2016), empirical corroboration remains limited due to a relative lack of diversity in the supporting evidence and the difficulty of deriving specific predictions from the theory about the precise timing of chills. Furthermore, the theory is ambiguous about whether chills occur when learning is required or when it is unnecessary (Pelowski et al., 2018), therefore making it unclear how to reconcile the theory with findings showing that MECs occur in response to unexpected musical events.

2.4 Discussion

We have conducted a systematic and critical review of the current literature on MECs, with the purpose of establishing a solid basis for future research. In this discussion, we first summarise each category of findings presented above, before integrating these findings in order to address the questions raised in the introduction. We then explore limitations of the reviewed research and of the present chapter, before providing a preliminary model of MECs and introducing a dataset listing pieces of music reported to elicit MECs in the reviewed literature. Finally, we outline a set of open issues, hypotheses, and recommendations for future research.

2.4.1 Summary of findings

Context

Most of the empirical work reviewed in Section 2.3.1 relies on the analysis of self-reports, and on a certain degree of subjective input from the researchers when it comes to interpreting and naming overarching categories and underlying factors. Taken together, however, these results suggest that, while MECs are a complex psychological construct, most of the population experiences them regularly, although not necessarily very frequently. MECs might comprise several psychologically distinct phenomena, are thought to be related to emotional and aesthetic experiences, and to involve a bodily sensation, which most often originates in the head, neck, back, or arms, and can include piloerection. Establishing a clear and consistent conceptual understanding of what exactly is being studied when researching MECs is a critical issue, and the research reviewed in that section provides a necessary first step in building such a framework.

Emotion and aesthetics

The relationship between MECs, emotions, and aesthetics is complex. The purpose of Section 2.3.2 was not to provide a comprehensive review of the literature on emotion and aesthetics, but rather to situate MECs within well-established frameworks of aesthetic and emotional responses to music, which are widely—though not always universally—accepted. From the evidence reviewed in that section, we conclude that MECs are a pleasurable psychophysiological response to music, and a possible, though not essential, component of emotional and aesthetic experiences of music. This makes them unsuitable as the sole indicator of such experiences, but if used in conjunction with self-reports, they provide attractive properties from an experimental point of view, because they are pleasurable, widespread, stable, memorable, discrete, and when accompanied by piloerection, objectively observable (Brattico & Pearce, 2013; Brattico & Varankaitė, 2019; Grewe et al., 2009b; McDermott, 2012; Sloboda, 1991; Stark et al., 2018; Vuust & Kringelbach, 2010; Zatorre, 2003).

Measures and paradigms

In Section 2.3.3, we found that self-reports and objective measures both provide distinct advantages, but also have their drawbacks. With self-report measures arise the issue of demand characteristics, through which the behaviour of participants can be influenced by the information they can infer about the experimental hypothesis (Juslin, 2016; Orne, 1962). Moreover, self-report measures are also subject to self-presentation biases and limited awareness of felt emotions, and providing them continuously or retrospectively can respectively cause issues with distraction or reliability (Zentner & Eerola, 2010). These problems do not arise with objective measures, but in the case of research on MECs, such measures are currently limited to the detection of piloerection, which does not encompass the entirety of the experience of MECs (see Section 2.3.1), potentially leading to increased type II error rates. Many studies have combined methodologies, thereby combining the complementary advantages of subjective and objective measures, and we advocate this approach in future research. In terms of paradigms, causal approaches have gained traction. They are crucial if we are to gain a better understanding of the causes of MECs, and should be used whenever possible in future research, along with naturalistic listening experiences to increase ecological validity (see Chabin et al., 2020; Eerola, 2018; Hargreaves & North, 2010; Hodges, 2016), longitudinal designs to study how experiences of MECs change over time (see Greasley & Lamont, 2016), and cross-cultural approaches to avoid an over-representation of classical music and Western participants, as is currently the case in research on MECs (see L. Harrison & Loui, 2014).

Physiological correlates

In Section 2.3.4, we found that MECs are associated with many physiological changes, and most often with increases in skin conductance response, heart rate, and respiration rate. However, for more ambiguous findings, the quality of the reviewed evidence must also be taken into account (Koelsch & Jäncke, 2015). Some studies systematically compared physiological responses in the presence or absence of MECs (e.g., Benedek & Kaernbach, 2011; Craig, 2005; Grewe et al., 2011; Grewe et al., 2009a; Guhn et al., 2007; Mas-Herrero et al., 2014; Mori & Iwanaga, 2017; Salimpoor et al., 2009; Sumpf et al., 2015), while other studies were correlational in nature, or compared averaged responses at the song level rather than continuous responses at precise moments in time (e.g., Balteș et al., 2011; Carr & Rickard, 2016; Jaimovich et al., 2013; Rickard, 2004; Salimpoor et al., 2011; Schäfer & Sedlmeier, 2011). Due to these differences in experimental design, greater weight should be given to findings about the presence of effects for skin conductance level and heart rate, and the absence of an effect for respiration rate. However, these studies are still limited by a lack of replication using different methodological approaches.

More generally, the reviewed evidence is consistent with increases in self-reported arousal when experiencing MECs (Balteș et al., 2011; Carr & Rickard, 2016; Grewe et al., 2009a; Mori & Iwanaga, 2015, 2017; Sumpf et al., 2015). However, the time course of physiological responses associated with MECs remains unclear, with changes in physiological arousal either preceding, co-occurring with, or following MECs, making it difficult to assess whether arousal is a cause or a consequence of MECs, or simply a co-occurring phenomenon. Physiological measures are sometimes thought to be relatively non-specific, and only indicative of a general state of arousal (Larsen et al., 2008; Panksepp & Bernatzky, 2002), but there is actually some degree of physiological response specificity, allowing particular response patterns to be associated with discrete emotional states (Hodges, 2016; Kreibig, 2010). Further research using a wider range of physiological responses could help identify which emotions are most closely related to MECs.

Neural correlates

In Section 2.3.5, we found that MECs involve the recruitment of brain structures associated with emotion, reward, pleasure, reinforcement, motivation, arousal, and motor processes (Blood & Zatorre, 2001; Brattico et al., 2009; Chanda & Levitin, 2013; Vuust & Kringelbach, 2010), and display activation patterns consistent with the reward experienced in response to food, sex, and drugs, notably through the involvement of dopaminergic and opioid systems (Blood & Zatorre, 2001; Chanda & Levitin, 2013; Mallik et al., 2017; Zatorre, 2003). Additionally, evidence suggests that individual differences in the experience of MECs might be due, in part, to differences in white matter connectivity between auditory and reward systems (Brattico, 2019; Hernández et al., 2019; Loui et al., 2017; Sachs et al., 2016).

Limitations of this body of work include poor generalisability due to small sample sizes and participants sometimes being selected for their ability to reliably experience MECs (e.g., Blood & Zatorre, 2001; Salimpoor et al., 2011), the poor time resolution of positron emission tomography, used in the study by Salimpoor et al. (2011), resulting in uncertainty about the precise timing of dopamine release (Habibi & Damasio, 2014; Vuust & Kringelbach, 2010), and finally, reliance on drawing reverse inferences about psychological mechanisms from observations of activation in brain areas subserving a broad range of psychological functions (Konečni, 2005; Logothetis, 2008; Poldrack, 2011). Overall, however, the consistency of the findings across a broad range of methods (neuroimaging, neurochemical, and lesion studies) provides strong support for the involvement of limbic and reward-related brain regions during MECs. A challenge for future research will be to understand if there are any patterns of neural activation which distinguish MECs from other instances of reward and pleasure.

Elicitors

Some authors have reported a lack of clear stimulus-response pattern with the experience of MECs (Bannister, 2020b; Grewe et al., 2007; Nagel et al., 2008), and while it is certainly true that a specific musical passage does not reliably cause MECs for all people (see Section 2.3.3), the evidence reviewed in the sixth section strongly points towards a set of acoustic, musical, and emotional elicitors being involved in the experience of MECs, including dynamic changes, increased roughness, crescendi, unexpected structural changes, textural changes, and perceived emotionality. Through underlying mechanisms, such as brain stem reflex and expectation, and associated factors, such as attention and familiarity, it is likely that, as is the case with aesthetic and emotional responses to music (Gabrielsson, 2011; Hargreaves, 2012; Juslin, 2013; Juslin & Västfjäll, 2008;

Scherer et al., 2001), MECs rely on an interaction between listener, context (about which there is currently relatively little research), and music. Importantly, most of the research discussed in Section 2.3.6 relies on correlational evidence, which weakens its strength. However, efforts have been made in recent research to use systematic manipulations in order to establish causality, confirming loudness and textural changes as elicitors of chills, for instance, and therefore resulting in a more robust understanding of the causes of MECs.

Individual differences

In Section 2.3.7, we found that in general, evidence for the influence of individual differences on the experience of MECs is mixed. This might be a consequence of most of these individual differences being studied in the context of exploratory research with little theoretical basis, with the exception of some of the Big Five personality traits (including openness to experience), as well as reward sensitivity, stylistic preference, and trait empathy. Regardless, from the totality of the evidence reviewed, it is now well established that openness to experience plays a role, and in more general terms, that personality differences affect who experiences MECs, and how often they are experienced. This should be taken into consideration when researching MECs, because individuals might react differently to various experimental situations, based on their personality characteristics.

Theoretical perspectives on function

At present, the theoretical accounts for MECs reviewed in Section 2.3.8 lag behind the empirical evidence in terms of their breadth, depth, degree of empirical corroboration, and ability to make clear and distinctive empirical predictions. Considering the diversity of empirical findings about MECs, it seems increasingly unlikely that a single functional mechanism could provide an adequate explanation for why they occur. There is currently little empirical evidence that specifically supports the separation call and knowledge instinct theories. Taken together, however, contrastive valence, peak arousal and pleasure, and the emotional state of being moved could account for much of the empirical evidence. It therefore seems plausible that competing theories based on evolutionary expectation and social processes might together explain the diversity in elicitors and personality characteristics involved in the experience of MECs (see Bannister, 2019).

It is worth emphasising again that all the theories on the function of MECs reviewed here are speculative, and would greatly benefit from the use of cross-cultural (see Beier et al., 2020), and developmental research, both of which (and ideally in combination) would provide evidence regarding the role of culturally-embedded learning in determining the elicitors and experience of MECs, as well as any potential evolutionary basis for their existence. In addition, hypothesis-based experiments testing the concrete predictions of the most promising theories are needed for further corroboration. In particular, the individual theories make different predictions about the psychological circumstances in which MECs would be experienced—during an experience of being moved (due to social closeness and empathy), during an experience of contrastive valence or auditory looming, or during an experience of high levels of emotional arousal or pleasure. Empirical experiments that test these predictions against one another are necessary to provide further clarity on the theoretical basis of MECs. Again, we would emphasise that it seems very possible that more than one psychological mechanism will be required to account for different kinds of chills.

2.4.2 Integration of findings

In this section, we integrate and expand upon the main findings from the reviewed literature in order to address the important questions raised in the introduction.

First, MECs seem to be relatively universal, as they are experienced by up to 90% of the surveyed population. However, while they are experienced with some degree of regularity, they remain a rare occurrence, with gaps between MECs sometimes reaching weeks or months for some people. This has no bearing on whether or not these experiences are meaningful to those who experience them, but it raises questions about their suitability for empirical research on emotional and aesthetic responses to music. Indeed, even though MECs are pleasurable, if they are rarely experienced, they are unlikely to provide a full picture of the aesthetic experience of music. Moreover, the fact that MECs are experienced by a smaller proportion of the population in experimental settings suggests that such settings may be inappropriate for the study of the entire range of aesthetic responses to music. However, the fact remains that MECs are pleasurable, and that despite their relatively sparse occurrence, they do reliably occur in experimental settings and are arguably the most convenient, objectively observable empirical measure of pleasure experienced in response to music listening. As long as care is taken not to place undue focus on MECs being representative of all aesthetic responses to music, we believe that they are worthy of scientific inquiry and have the potential to reveal much about how music is processed and about music appreciation.

Second, we argued that while MECs exhibit some characteristics consistent with emotional responses, the lack of clear motor expression and action tendency suggests that they should not be considered as an emotional state per se. This argument could reasonably be debated. Brain structures associated with motor processes are recruited during MECs, possibly suggesting preparedness for singing and dancing (Brattico & Pearce, 2013), and to our knowledge, there has been no investigation of facial muscle activation in MECs. Moreover, a possible action tendency exhibited by MECs could consist of a focus of attention towards aesthetic stimuli. It could therefore be argued that MECs are indeed an emotional reaction, or at least, the manifestation of an extreme emotional reaction. While we are inclined to consider MECs as a psychophysiological response which can form part of a range of emotional states, further research is needed to establish the precise relationship between MECs and emotions. Regarding the relationship between MECs and aesthetic responses, the evidence strongly suggests that MECs are perceived as pleasurable. While there are reports of displeasurable chills, they tend to occur in reaction to non-musical stimuli, and we therefore surmise that most, if not all, MECs are pleasurable, possibly because music listening itself is generally a pleasurable activity. However, MECs are often used as an empirical indicator of peak pleasure, despite there being relatively little evidence in support of this claim. In our opinion, conflating MECs and peak pleasure misrepresents the relevance of MECs to research on music emotion and aesthetics. We believe that research on MECs contributes usefully to the literature on emotional and aesthetic responses to music, as long as MECs are not considered as fully fledged emotional and aesthetic responses, but rather as non-obligatory but enhancing components of these responses.

Third, we reviewed two articles which suggest that chills represent a collection of phenomenologically and psychologically distinct experiences. These are, to our knowledge, the only contributions studying this question from an empirical point of view. However, it is worth emphasising that in both cases, several stimulus modalities were investigated, as opposed to music only, and it is not yet known whether or not MECs specifically are a multi-faceted phenomenon as well. We believe it is likely that MECs can arise from different combinations of elicitors and associated underlying mechanisms, which could be explained by different theories on the function of MECs, but it remains to be determined whether these would lead to different types of MECs, or to the same psychophysiological response in all cases. The implications for past and future research on MECs are considerable, because failing to distinguish between different types of MECs could lead to null, conflicting, or misleading empirical findings. We therefore believe that investigating this question should be a priority for research on MECs.

Fourth, the evidence suggests that there are great similarities between the physiological and neural correlates of MECs and pleasurable responses to music in the absence of MECs (see Archie et al., 2013; Koelsch, 2010, 2014; Salimpoor et al., 2011), and that so far, there does not seem to be any physiological or neural signature setting these two responses apart. This is due to the fact that, as discussed above, MECs and pleasure are tightly coupled. There have been no studies investigating this specific issue, so for the moment being, the only available evidence relies on drawing inferences from tangential findings. In a lesion study (Grunkina et al., 2017) and in a study about anhedonia (Mas-Herrero et al., 2014), there have been reports of participants experiencing MECs with an impaired ability to experience pleasure, which would suggest that these two responses are not always associated. However, in both cases, it is not possible to establish if these participants had the ability to experience some residual degree of pleasure. It is therefore not currently possible to establish whether MECs invoke general-purpose mechanisms involved in emotional processing and reward, or dedicated neural and physiological machinery instead. We suspect that there are specific neural and physiological signatures for MECs, but that they will only be uncovered in a systematic empirical comparison between MECs and other highly pleasurable experiences of music.

Finally, while we have already discussed causes of MECs in this chapter, it is worth integrating the evidence on their elicitors, individual differences, and origins. The evidence suggests that while MECs are idiosyncratic to some extent, stimulus-driven characteristics, such as changes in loudness, crescendi, or emotionality, strongly drive their occurrence. In terms of individual differences, there was no clear effect of age, gender, or musical training, but we believe individual differences should not be dismissed just yet. If expectation is involved in MECs, there could be an effect of musical training, since musicians tend to develop more precise expectations (Hansen & Pearce, 2014; Hansen et al., 2016; Quiroga-Martinez et al., 2019), or an effect of individual differences in music perception abilities. Openness to experience was identified as strongly associated with MECs. This personality trait is associated with preference for sophisticated, intense, and mellow music (Schäfer & Mehlhorn, 2017), which could imply that individuals with high scores on openness to experience seek out music which features violations of expectation or high emotionality. In terms of theoretical accounts of the function of MECs, the approaches most consistent with the available evidence are that MECs occur because of the emotional state of being moved, peak arousal, or contrastive valence. Further research is needed to test these theories against one another. However, it is also possible that all these theories account for different aspects of MECs. As discussed above, we speculate that, regardless of whether there exist different types of MECs or not, it is possible that there are different causes of MECs. For instance, acoustic elicitors could engage the mechanism of brain stem reflex, causing MECs through peak arousal, the function of which is perhaps associated to a need to maintain homeostatic balance. Musical elicitors could engage musical expectation, causing

MECs through the evolutionary fear-based process of contrastive valence. And finally, emotional elicitors could engage the mechanism of emotional contagion, causing MECs through the evolutionary social process of being moved. This would signify that MECs rely on an interaction between listener, context, and music, which is itself driven by psychological mechanisms of emotional responses to music and evolutionary reasons for the function of MECs.

2.4.3 Limitations

In this section, we provide a broad overview of the quality of the reviewed research, as well as a discussion of the limitations of the present chapter.

Due to the breadth of topics (see Section 2.2) and wide range of study designs (see Table 2.2) in the research covered in this chapter, it is difficult to provide an effective summary of research quality. It is however possible to highlight common pitfalls and notable exceptions in terms of methodological issues. First, there is the issue of validity and reliability of measures of MECs. As noted earlier, there is currently a trade-off between validity and reliability. When tested, the Goosecam accurately identified each occurrence of piloerection (Benedek et al., 2010), and can therefore be considered as reliable, but since MECs do not consist of piloerection only, it could be argued that the Goosecam and other objective measures are not comprehensively valid measures of MECs. Inversely, self-reported measures are inherently more valid (although this depends to some extent on the definition of MECs provided to the participants in a given study), but they are less reliable due to potential interference from demand characteristics, self-presentation biases, limited awareness of felt emotions, distraction, or the retrospective nature of self-reports. This leads to limitations in the present chapter, since it is difficult to assess if it is indeed MECs that were investigated in all of the reviewed research, and if all MECs were reported by participants in said research. Issues of validity and reliability are crucial to research on MECs, and while combining objective measures and self-reports can address some of these issues for the time being, it is hoped that future research will identify better measures of MECs.

Then, there is the issue of methodological quality and risks of bias. Despite a long history, research on MECs is still in its exploratory stage. There is currently a lack of underlying theoretical framework for hypothesis-based research, leading to large disparities in terms of research methods, and therefore preventing their implementation details (and by extension, their risks of bias) from being systematically compared across studies. This lack of consistency is also what makes a meta-analysis of some facets of research on MECs currently difficult, if not impossible. For instance, in the first sections of the present chapter, many qualitative studies were discussed. On the one hand, broad preliminary investigations of emotional responses to music featured MECs as a component reported by some participants (e.g., Panzarella, 1980; Scherer et al., 2001; Sloboda, 1991), and were often characterised by the recruitment of non-representative convenience samples, high attrition rates, and coding of qualitative data by a single researcher. On the other hand, purposeful investigations of chills as a psychological construct (e.g., Bannister, 2020b; Maruskin et al., 2012) benefited from exhaustive and transparent methodologies, attempts to recruit large and more representative samples of the population, and explicit investigations of inter-rater reliability. Similarly, in the reviewed quantitative research, methods diverged widely, with the majority consisting of some form of cross-sectional study (see Table 2.2). There were a few double-blind, randomised studies (e.g., Ferreri et al., 2019; Goldstein, 1980) with very large disparities in terms of transparency of research methods and statistical analyses, resulting in limited reproducibility for Goldstein's (1980) study. In many cases, participants were exposed to many conditions, and while single blinding was attempted in most cases and is certainly possible to some extent, it is likely that the participants could infer the hypothesis at hand. For instance, if participants are asked to report MECs in response to music they brought with them as opposed to unfamiliar music, it is reasonable to assume that the participants might have guessed that they should experience more MECs when listening to their own music. This lack of effective single blinding, along with the absence of researcher blinding in most cases, most likely resulted in some degree of response and experimenter bias in many of the reviewed studies. In addition, many quantitative studies involved the participation of undergraduate populations with an over-representation of musically trained individuals, were not clearly hypothesis-driven, and did not feature power calculations or report effect sizes. The combination of these factors affects the quality of the evidence in the present chapter, making it difficult to assess the relative strength of the reviewed evidence and the extent of the presence of specific biases. While there is a distinct trend favouring more controlled experimental approaches in recent years, much remains to be done in terms of adopting best scientific practices. This is expected to some extent from an area of study which is still in the process of establishing a theory-grounded research agenda, but it is hoped that the present thesis will provide substantial impetus in that direction as research on MECs matures.

Finally, in addition to the limitations already presented in this section, the methodology of the present chapter also suffered from some limitations. First, while contributions in languages other than English were considered for inclusion in this review, databases were only searched using English terms. These search terms did yield some research written in other languages, and we are not aware of any additional such literature, so we suspect that it would not be extensive. If any does exist, it could have influenced the findings of the present review to some extent, perhaps due to different stylistic enculturation leading to different expectations in participants in such research, or to different perceptions of meaning in pieces of music, for instance. We suspect, however, that such differences would most likely be small if not completely absent, since we already know that people from different cultures can experience MECs (Beier et al., 2020), and that music from many genres and countries of origin can cause MECs (see dataset in Section 2.4.4). Second, while publication bias is most likely present, it is difficult to assess its extent in research on MECs due to the relative lack of availability of unpublished research on the topic. While meta-analytic methods such as funnel plots could help investigate publication bias, they were not suitable for the present chapter due to the need to encompass the great diversity of topics and methods in research on MECs. Publication bias increases the risk of drawing false conclusions (type I errors in particular), but we estimate that risk to be limited in this case, because most of the reviewed research is exploratory, therefore leading to the reporting of many positive and negative findings within individual papers. However, future research should seek to limit publication bias by using more specific definitions of MECs, testing fewer but more hypothesis-driven relationships, pre-registering research protocols, conducting better-powered studies, and harmonising research designs across studies. Third, no formal coding of relevant research insights was conducted due to the variety and complexity of empirical findings, opening the possibility for some degree of researcher bias, and therefore increasing the risk of drawing false conclusions. Overall, however, we believe these limitations were necessary in order to allow broad and integrative coverage of empirical and theoretical research on MECs. As a result, the risks of bias are balanced by the fact that we do not consider the outcome of the present review a finalised theory of MECs, but rather a framework for future hypothesis-driven research.

2.4.4 Framework for future research

We conclude this chapter by providing a tentative, preliminary framework for future research on MECs. We begin with a set of minimum criteria for a response to music to be considered as MECs. We then provide a model of MECs based on the reviewed literature and a dataset of pieces of music known to cause MECs, before delineating open issues, hypotheses, and recommended approaches.

Criteria for MECs

We provide here a preliminary set of minimum criteria for MECs. These criteria rely, in part, on our current interpretation of the strength of the available evidence, and will certainly be subject to change as research on MECs progresses. They are also conservative, only including criteria which the evidence suggests are almost certainly associated with MECs, but not criteria based on other findings which have yet to produce consensus.

We argue that, for a typical individual's response to music to be categorised as MECs, this response should be a sudden, fleeting, and pleasurable psychophysiological reaction to music-driven properties (whether they are acoustic, musical, or emotional), most commonly originating from the head, neck, back, or arms. Its occurrence should be possible in a large proportion of the population, and particularly in individuals with high openness to experience. It should involve increases in subjective arousal, skin conductance response, and heart rate, as well as limbic and reward-related neural activity. This response might but does not necessarily involve piloerection, and forms part of an emotional or aesthetic experience of music, though emotional or aesthetic experiences of music need not necessarily involve this response.

Model of MECs

The criteria presented above are integrated with the rest of the literature in a preliminary model of the experience of MECs, in order to provide a framework which will allow the formulation of hypotheses for future research on MECs. This model is not exhaustive, but it includes a range of parameters, mechanisms, and response attributes we believe to be the most relevant to future investigation of MECs.

The model is presented in Figure 2.2. Parameters represent the interaction between listener, music, and context that is most likely involved in MECs. This aggregation of parameters gives rise to the response of MECs, through the combination of the psychological and evolutionary mechanisms we identified as the most likely to underlie MECs. We included elements which lack full empirical verification, but which, in our view, represent important open issues in research on MECs, such as the effect of attention, the exact nature of the psychological mechanisms which lead to MECs, or the extent of the relationship between aesthetic and emotional responses and MECs. A distinguishing feature of this model is that it also groups phenomena which we believe could be related across categories, and could provide different pathways for the experience of MECs, if not different types of MECs. For instance, it could be that individuals with high trait empathy, perceiving emotional elicitors when listening to music,

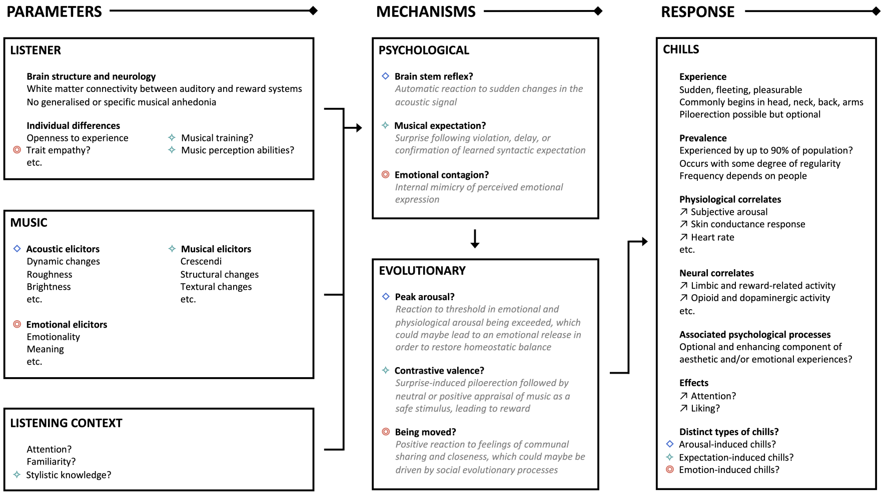

Figure 2.2: Preliminary model of MECs. Parameters on the left represent factors which influence the response of MECs on the right, via psychological and evolutionary mechanisms in the middle. Diagonal arrows represent increases in the associated response. Sentences in italics represent definitions for the listed mechanisms. Question marks represent open questions which lack empirical corroboration. The term “etc.” indicates categories for which future evidence or replication of current evidence may warrant the addition of further entries. Symbols link together phenomena which could be related, and could contribute to distinct experiences of MECs.

mimic the perceived emotion through emotional contagion, leading to MECs (or even to a distinct type of emotion-induced MECs) through the process of being moved. While such predictions are preliminary, they integrate existing findings, highlight important open issues, and allow the formulation of new hypotheses, providing a path towards a better understanding of MECs.

Dataset of MECs

Empirical studies of music-evoked emotions most often feature stimuli pertaining to MECs (Warrenburg, 2020), such that a large quantity of music which can cause MECs has been documented in the academic literature on MECs. With the aim of facilitating more integrated research on MECs, we have compiled *Chills in Music (ChiM)*, a dataset which contains, to our knowledge, all pieces of music which have been reported to elicit MECs in the literature reviewed in the present chapter[2]. Details about the preparation of ChiM are available in Chapter 4. It should be noted that the dataset contains little information about the timing of MECs in most pieces of music, due to limited information in the

[2]ChiM is available at https://doi.org/10.17605/osf.io/uyg7m.

reviewed literature. Efforts should be expanded to augment ChiM with precise timing information, in order to support future computational research on MECs.

Open issues and recommendations

In this section, we highlight open issues in the literature on MECs, based on the reviewed literature and on the preliminary model presented above. In our view, investigating these issues has the most potential to advance research on MECs. Throughout this systematic review, we also identified significant methodological shortcomings regarding research design, adequacy of experimental variables, measures of MECs, and terminology. We provide suggestions for addressing these shortcomings below.

Table 2.7 lists what we consider to be the most important open issues in the available evidence on MECs, along with hypotheses and recommended experimental approaches. While all these issues are derived from the model of MECs we provided, we make a distinction between issues arising from the reviewed literature, and specific predictions arising from the proposed model.

Throughout this chapter, we have provided methodological recommendations to address shortcomings in the research on MECs. Notably, we recommended that piloerection should not be used as the sole indicator of MECs, that MECs should not be used as the sole indicator of emotional and aesthetic responses, and that individual differences should be taken into account, particularly because chills could be a multi-faceted phenomenon, which could lead to null, conflicting, or misleading results if this is not taken into consideration. We argued that a combination of self-reports and objective measures are currently best suited for the study of MECs, and that care should be taken when validating self-reports of MECs with skin conductance response. Finally, we recommended the use of the terms chills and piloerection, and suggest a definition for participants in research on MECs, characterising MECs as a fleeting, pleasurable bodily sensation, sometimes accompanied by goosebumps, experienced when listening to specific musical passages.

2.4.5 Conclusion

We conducted a systematic review of the literature on MECs. Theoretical and empirical findings were integrated, leading to the conclusion that MECs are a prevalent psychophysiological response which can include piloerection, and a pleasurable, though not essential, component of emotional and aesthetic experiences. They have been studied using both subjective and objective measures, with a recent focus on causal approaches—a necessary endeavour due to most of the evidence being correlational in nature, and therefore often difficult to

Table 2.7: Open issues, hypotheses, and suggested approaches

Open issue	Hypothesis	Suggested approach
Universality of chills	MECs are experienced by the same proportion of the population, regardless of culture, but are dependent on enculturation.	Conduct a large-scale, cross-cultural survey of MECs, recording information about exposure to various musical cultures and genres.
Occurrence of piloerection	Piloerection occurs once MECs exceed an intensity threshold.	Record self-reported intensity of MECs and compare to measured piloerection. This might require further validation or development of piloerection sensors.
Specificity of MECs	MECs exhibit different physiological and neural signatures than those of emotion or pleasure.	Compare responses with self-selected music that can elicit distinct experiences of MECs, emotion, and pleasure. If there is specificity, a classifier could be trained to distinguish unlabelled instances of MECs from emotional and pleasurable episodes without MECs.
Acoustic and musical elicitors	The effect of acoustic elicitors on MECs is partially mediated by musical elicitors, and vice versa.	Compare extracted acoustic and musical features (using music information retrieval and/or manual annotation) around the onset of MECs (using a dataset such as the one provided in this review). Alternatively, systematically manipulate stimuli to independently vary the two types of elicitors and compare occurrences of MECs.
Familiarity	MECs are experienced more frequently as familiarity increases.	Use a longitudinal design to study the progress of the frequency of MECs when repeatedly exposed to previously unfamiliar and familiar music with the potential to elicit MECs.
MECs and attention	Attending to music increases the likelihood of MECs occurring, and MECs focus attention towards the eliciting music.	Assess the occurrence of MECs at rest and during a non-musical distractor task while listening to music. Fewer MECs should occur while distracted, and if they occur, they should impair performance on the task.
Psychological mechanisms	Exploratory animal and developmental research can help pinpoint the psychological and neural mechanisms underlying MECs.	Since brain stem reflex, musical expectation, and emotional contagion rely on different psychological and neural mechanisms, which might be more or less well developed in different species and at different developmental stages, exploring the prevalence of MECs in animals and individuals varying in developmental age could shed light on the mechanisms underlying MECs, and identify developmental trajectories.
Peak arousal	MECs can occur in response to peaks in arousal or pleasure, but might not always since several mechanisms drive the occurrence of MECs.	Record measures of physiological arousal for a large number of MECs to identify a threshold for peak arousal or subjective pleasure. MECs should happen every time this threshold is exceeded, but could also happen below this threshold if elicited by a different mechanism.
Musical expectation	MECs can occur in response to violations of expectation, but might not always since several mechanisms drive the occurrence of MECs.	Collect precise timing information for when MECs occur (or use the dataset provided in this review), and compare them to the output from a computational model of expectation. MECs should always occur for sufficiently strong violations of expectation but might occur elsewhere if elicited by a different mechanism.
Evolutionary mechanisms*	MECs can occur via either peak arousal, contrastive valence, or the process of being moved.	Carefully prepare stimuli with the potential to elicit MECs via these three mechanisms, controlling for the others, and collect continuous measures (for instance, the two measures detailed above for peak arousal and expectation, and self-reports for being moved). Peaks for each measure should correspond to the onset of MECs for each targeted mechanism.
Listener and context*	Susceptibility to MECs caused via different mechanisms is partly governed by individual differences, familiarity, and stylistic knowledge.	Using the approach detailed above, compare individual differences and personality correlates across participants who reported the most MECs for each mechanism. Include familiarity and stylistic knowledge for each piece of music as a random effect in a mixed effect model.
Distinct types of MECs*	Different parameters and mechanisms cause different types of MECs, with distinct physiological and neural signatures.	Similarly, using the approach detailed above, compare physiological and neural correlates for MECs elicited via each mechanism. Alternatively, collect these measures along with qualitative descriptions of MECs to identify differences between different categories of MECs.

Note. * Predictions derived from preliminary model of MECs.

interpret. In terms of biological basis, MECs are associated with physiological changes and increased arousal, and recruit brain structures and systems relevant to emotion, reward, and motivation. We reviewed many possible causes of MECs in this chapter. In light of the quality and quantity of the evidence, we believe certain factors to be of particular importance. Notably, MECs can be elicited by acoustic, musical, and emotional stimulus-driven properties which, taken together, suggest a prominent role of sudden changes in acoustic properties, of high-level structural prediction, and of emotionality. They are influenced by personality differences, and especially openness to experience, which is a strong predictor of the ability to experience MECs. Finally, the more convincing theoretical accounts of the function of MECs suggest an involvement of mechanisms based on expectation, peak emotion, and being moved.

We concluded this chapter by establishing a preliminary framework for future research on MECs, providing a set of minimum criteria for a response to music to be considered as an instance of MECs, a model of MECs that explicitly allows for different psychological pathways for the experience of MECs and different types of MECs, a dataset of pieces of music known to cause MECs, and a list of open issues, hypotheses, and potential experiment approaches.

Chapter 3

Preference and familiarity

In Chapter 2, we went over methodological limitations in previous research on MECs, and outlined a set open issues based on the reviewed literature. The present chapter introduces a study following some of the recommendations laid out in Chapter 2, in which the relationships between MECs, piloerection, pleasure, musical content, stylistic preference, familiarity, and liking are examined in a controlled, longitudinal experiment.

3.1 Introduction

Broadly speaking, the present study aimed to investigate the effects of three independent variables (musical content, stylistic preference, and familiarity) on three dependent variables (MECs, piloerection, and pleasure), as well as the extent to which static ratings of liking for pieces of music could be predicted by all of these factors combined.

Musical content here refers to the collection of stimulus-driven characteristics that might influence the propensity of a specific piece of music to induce MECs. As discussed in Chapter 2, MECs have previously been considered to be highly idiosyncratic. There is merit to this claim, given that in many studies, one participant's MECs-inducing stimulus is successfully used as another participant's control stimulus. However, this view is incompatible with the fact that there is evidence (though mostly correlational in nature) for an association between MECs and a range of clearly defined acoustic and musical elicitors, which should logically lead to some pieces of music being more likely to induce MECs than others. It is unlikely that a specific combination of elicitors always induces MECs for everyone, or conversely, that MECs can be experienced when listening to any piece of music. Instead, we would expect that, in aggregate, pieces of music which induce MECs for some people are more likely to induce MECs in others than randomly selected pieces of music.

A possible way to establish causation for this claim is to assess the occurrence of MECs when listening to tracks that have previously been reported to elicit MECs compared to tracks that have not. The possibility of confounding factors can be limited by matching this second set of tracks with the first as closely as possible on every parameter (including, in this case, artist, duration, and popularity) except musical content, resulting in two comparable pools of tracks (hereafter referred to as *sources*). While it is likely that some tracks from the matched source also have the ability to cause MECs, we would expect tracks from the chills source (i.e., tracks previously reported to elicit MECs) to cause more MECs (see Chapter 4 for a more thorough discussion of the assumptions behind this study design). When it comes to musical content, the hypothesis was therefore that tracks which cause MECs for some people are more likely to cause MECs for other people. Causal evidence from this study should help clarify if this is the case, which would suggest that there is a clear effect of acoustic and musical elicitors on the occurrence of MECs, or if MECs can be experienced when listening to any music, which would suggest that they are an indicator of individualised responses to music, or at least that other factors are involved.

Stylistic preference and *familiarity* are two of these factors (included in the *listening context* section of the model of MECs introduced in Chapter 2), and are likely to be involved either way, considering that the occurrence of MECs is not determined by stimulus-driven properties alone. Previous research identified conflicting effects of stylistic preference (Bannister & Eerola, 2018; Nusbaum & Silvia, 2011), but these findings were limited in two ways. First, the hypothesis was that the occurrence of MECs might be driven by preference for a specific subset of musical genres, resulting in findings that individuals who experience MECs tend to prefer reflective and complex genres in one study (Bannister & Eerola, 2018), or upbeat and conventional genres in the other (Nusbaum & Silvia, 2011) as measured by STOMP, the Short Test Of Music Preferences (Rentfrow & Gosling, 2003). However, if MECs involve an interaction between listener, context, and music (see Chapter 2), we would expect an individual's stylistic preference for any genre to have an effect on the occurrence of MECs. In other words, rather than the characteristics of specific genres leading to more MECs, stylistic preference should drive the occurrence of MECs by making them more likely to occur in an individual's preferred genre, through a combination of personal taste and stylistic enculturation, possibly linked with an effect of expectation on MECs (see Beier et al., 2020).

Second, there are limitations to the methods that were used to assess stylistic preference. STOMP was designed to investigate personality differences in the preference for specific musical genres, and consists of asking participants to provide Likert scale preference ratings for 14 genres. More recent approaches require participants to rate their preference for musical exemplars assigned to each genre, instead of rating labelled genres directly (Bonneville-Roussy et al., 2017; Rentfrow et al., 2011). Such approaches are limited because they all rely of some degree of reductionism and subjective interpretation in genre labelling and categorisation (by participants or researchers), which is therefore not consistent across individuals, and because most adults have omnivorous stylistic preferences (for a review, see Greasley & Lamont, 2016). For the purpose of the present study, this issue was circumvented by directly asking participants whether a given track was in a liked genre or not, without relying on labelling said genre, allowing for an objective and causal investigation of the hypothesis that more MECs are experienced when listening to tracks in preferred musical genres.

Familiarity, similarly to stylistic preference, could possibly affect the occurrence of MECs. There is a long history of research on the link between familiarity and liking, including seminal work such as Zajonc's (1968) mere exposure effect and Berlyne's (1971) inverted-U relationship. Overall, previous research identified a strong relationship between familiarity and liking for music (for a review, see Greasley & Lamont, 2016). Of relevance to research on MECs, familiarity was even identified as more important than liking for emotional engagement with music (Pereira et al., 2011). When it comes to direct empirical evidence of an association between familiarity and MECs, however, results are mixed and were affected by methodological limitations, as discussed in Chapter 2. Notably, repeated exposure was only assessed in the context of a single experimental session (Balteș et al., 2011; Bannister, 2020c; Blood & Zatorre, 2001) or over a longer period, but for a single participant (Grewe et al., 2007), or the effect of familiarity was only assessed statically, by comparing ratings of familiarity across stimuli (see Chapter 2). Longitudinal methods have been underused in research on familiarity and liking (Greasley & Lamont, 2016), but they remain the best way to systematically manipulate familiarity in order to show causal effects on MECs. The present study used such an approach, with the hypothesis that increased exposure, and therefore familiarity, results in increased occurrences of MECs.

In terms of dependent variables, the present study examined all of MECs, piloerection, and pleasure. The reasons for considering different dependent variables instead of MECs only were threefold, and are all discussed in Chapter 2. First, not all MECs are accompanied by piloerection, so there is a need to investigate whether MECs and piloerection are similarly affected by changes in track source, stylistic preference, and familiarity. Second, while there is a documented association between MECs and pleasure, there have been few attempts, if any, to disambiguate the two responses in previous research. Finally, while the use of objective measures would have been preferred, such as the automatic detection of piloerection, we have argued that a combination of self-reports and objective measures is currently best suited for the study of MECs. Collecting data on these three variables therefore allowed us to explore the hypotheses that piloerection overlaps with self-reported MECs, and that MECs are more likely to occur in musical passages perceived as pleasurable.

The research design for this study aimed to provide ways to systematically manipulate all three independent variables, while also following the methodological recommendations laid out in Chapter 2, which argue for the use of naturalistic listening experiences using existing pieces of music in order to increase ecological validity, which are both considered crucial when investigating aesthetic and emotional responses to music (Eerola, 2018; Hargreaves & North, 2010; Hodges, 2016). In summary, the main hypotheses underlying this study were that the occurrence of MECs increases for tracks previously identified to cause MECs (as discussed above), with stylistic preference, and with familiarity (both included as hypotheses of interest in Chapter 2), that there is an overlap between piloerection, MECs, and pleasure in music (as discussed in the previous paragraph), and that the combination of all of these factors has an effect on overall liking for pieces of music (a key motivation for the present thesis).

3.2 Methods

3.2.1 Stimulus selection

Online survey

We conducted a survey study hosted on Qualtrics (Qualtrics, Provo, UT) with several distinct objectives: 1) build a set of ecologically valid stimuli from a single, controlled, and contemporary source for the purpose of the present study, 2) build a large dataset of onset of MECs for the computational analysis detailed in Chapter 5, 3) collect some basic demographic information for further investigation into the effects of individual differences on the occurrence of MECs, and 4) contribute to building a pool of potential participants for the present study. This survey was conducted online in order to access a wider and more representative sample of the population.

Survey responders were asked to report basic demographic information such as age, gender, and country of residence. They were then asked to provide up to ten pieces of music that include instants at which they often experience chills (defined here as shivers, goosebumps, or a tingling sensation experienced in response to music listening). For each piece of music, participants were asked to provide the name of the artist or composer, the title of the piece of music, a link where the piece of music can be streamed online (on Youtube, Spotify, SoundCloud, or a similar platform), and at least one and up to five precise timestamps for instants at which they often experience the onset of chills (i.e., the exact moment at which chills tend to begin). Participants were advised not to include music which can be strongly associated with specific personal memories, such as life events (e.g., a wedding, a memorable concert, etc.) or periods of time (e.g., summer of 2013, secondary school, etc.), or music that was taken from a film soundtrack if they had watched the film in question. These restrictions, taken from a previous study by Salimpoor et al. (2009), were meant to maximise the possibilities that the MECs in question were induced by widespread and detectable acoustic and musical elicitors, as opposed to autobiographical elicitors driven by episodic memory, which would not generalise across participants.

Survey responses for the present study were collected between February 2018 and May 2018, although the survey was left running for longer to collect more onsets of MECs for the computational study presented in Chapter 5. The questionnaire was disseminated on a wide range of online platforms, including international academic mailing lists, staff and student mailing lists within Queen Mary University of London, as well as Twitter[1] and Reddit.[2] Complete responses were collected from 221 participants, ranging in age from 18 to 77 years (M = 25.3 years, SD = 9.4 years). Of the participants who reported their gender, 72 identified as female and 144 as male. Responses originated from a wide range of geographical areas (50 % North America, 37% Europe, 5% Asia, 5% Oceania, 2% South America, 1% Africa). The resulting data required some manual cleaning, consisting of removing entries which were not changed from the default answers provided in the survey, as well as discarding non-valid URLs. This process resulted in retaining 214 complete responses, corresponding to 671 tracks.

From these 671 tracks, we needed to subset a pool of suitable stimuli according to the requirements of the study. This process involved four steps. First, tracks which were not available for streaming on Spotify were discarded. Second, tracks which were longer than five minutes in duration were discarded (with a few exceptions discussed later), in order to keep the duration of the experiment manageable, considering that the participants would need to listen to several tracks many times throughout the lab experiments and the longitudinal phase of the study. This duration is consistent with similar studies requiring participants to listen to several tracks in one sitting (Laeng et al., 2016; Mori & Iwanaga, 2014b), and was preferred to selecting excerpts from each track (Blood & Zatorre, 2001; Sachs et al., 2016; Salimpoor et al., 2009), so as to ensure ecological validity by allowing participants to listen to tracks in full. Third, tracks for which MECs

[1]https://www.twitter.com
[2]https://www.reddit.com

were reported within ten seconds of the beginning or the end of the track were discarded. This threshold was set arbitrarily, with the objective to maximise the chances of new listeners experiencing MECs. Finally, tracks for which MECs were reported within ten seconds of any passage including sung lyrics were discarded. It was decided to implement this restriction because sung lyrics have the potential to introduce additional confounding factors. This is often accounted for in previous research by asking participants to provide instrumental music only, but we felt this was overly restrictive for the purpose of the present study. These thresholds (five minutes for track duration and ten seconds for buffers between reported MECs and track beginning, end, and sung lyrics) were chosen by comparing different threshold values in order to retain an adequate number of tracks while minimising track duration and maximising buffers.

This process results in a pool of 93 tracks, which we subsequently refer to as tracks from the *chills source*.

Matching procedure

As discussed above, a part of the experimental design for this study relied on comparing tracks previously reported to cause MECs to tracks that were not, matched as closely as possible on every parameter except musical content, resulting in two comparable pools of tracks. The rationale behind this manipulation was that, considering that acoustic and musical elicitors of MECs have been identified in previous research, it logically follows that some pieces of music should include more of these elicitors than others. If the hypothesis that, as a consequence, music which induces MECs for some people is more likely to induce MECs in others, tracks from the chills source should then be more likely to induce MECs than tracks from the matched source. This reasoning is central to several experiments in the present thesis, and the assumptions and limitations behind it are covered in more detail in Chapter 4.

The matching procedure used in this study was inspired by the one used by Jakubowski et al. (2017). In their study, the authors aimed to obtain close matches for a set of 100 tracks that were named as causing involuntary musical imagery (otherwise known as earworms) in order to run statistical comparisons between both sets of tracks. Matching was performed by identifying a pool of candidate tracks with similar artists, time periods, chart positions, and genres as the target set of tracks. Matching was then conducted algorithmically, allowing precise matching on several variables at once: artist, genre, and chart information (highest entry, longevity, and days since exit).

While the procedure for the present study was conducted manually, a similar algorithmic procedure is detailed in Chapter 4. In the present study, each track from the chills source received five candidates for matches, which were manually shortlisted in the following way. First, a match should not have been mentioned in any of the survey responses as a track which can elicit MECs. Second, a match should be from the same artist as the target track. Third, its duration should be between 2.5 and 5 minutes. Fourth, its popularity, as assessed by the number of plays on Spotify, should be between half and double that of the target track. Finally, when possible, it should be sourced from the same album as the target track. If not, search should be expanded to tracks produced as close as possible in time to the target track. This process allowed to select candidate tracks while controlling as much as possible for stylistic preference (following the assumption that artists tend to produce tracks in relatively closely related genres), duration (in order to allow for similar opportunities to experience MECs), and quality (using popularity as a proxy for this more abstract measure). Each candidate track was ranked manually, resulting in subjective rankings of how closely related they were to their target track.

The top three matches for each track were retained, resulting in a pool of 279 tracks, which we subsequently refer to as tracks from the *matched source*.

3.2.2 Software and device

Stimulus allocation

The objective of the stimulus allocation online session was to select an individual stimulus set of 20 unfamiliar tracks for each participant, consisting of five tracks for each combination of source (chills or matched) and stylistic preference (liked or disliked). This set of tracks would then be used by this participant for the rest of the study.

Unfamiliarity was key to the present study, so there was a need to limit exposure to each track as much as possible before the first lab session. In order to do so, short excerpts were extracted for each track. The duration of said excerpts was decided based on previous research. While the valence of a piece of music can generally be recognised in excerpts lasting as little as one-eighth of a second (for a review, see Mace et al., 2012), it generally takes up to a second to identify the genre of a piece of music (Gjerdingen & Perrott, 2008; Mace et al., 2012). When it comes to familiarity or song recognition, however, estimates vary between less than a second and up to ten seconds (Jensenius, 2002; Krumhansl, 2010; Pereira et al., 2011; Schellenberg et al., 1999), with the added caveat that salient features are required (Jensenius, 2002), such as dynamic, high-frequency spectral information (Schellenberg et al., 1999). To minimise the likelihood of familiar tracks being selected for the first lab session, we opted for an excerpt duration of 15 seconds, which we deemed short enough to not affect the outcomes of the present study while safely allowing participants to recognise both familiarity and stylistic preference for a given track.

There is much less information available about which excerpt should be sampled from all possible excerpts present within a track, with Mace et al. (2012) reporting random samples taken between 00:30 and 04:00 for each track, and other authors simply not reporting that information (e.g., Bonneville-Roussy et al., 2017; Krumhansl, 2010; Pereira et al., 2011). Randomly selecting excerpts felt inadequate, so we opted for the heuristic of choosing the loudest moment for each track as long as it didn't correspond to a previously reported onset of MECs, following previous recommendation about salience and dynamics (Jensenius, 2002; Schellenberg et al., 1999), and assuming that this would often correspond to the most recognisable moment of a track. More precisely, each track was trimmed by 10% in duration on each side, and a moving average of absolute amplitude was computed using a 15-second sliding window. A 15-second excerpt centred around the peak of the moving average was selected for each track, as long as no onsets of MECs were reported during or within five seconds of the excerpt. If MECs were previously reported near the excerpt, the process was repeated using each following peak in average amplitude until a suitable excerpt was identified for each track. A one-second fade in and fade out was applied to each excerpt, before applying Root Mean Square (RMS) normalisation to harmonise loudness between excerpts, and exporting the excerpts to MP3 format at 170–210 kbps to reduce loading times during the online session. Audio computations and manipulations were conducted using the *tuneR* (Ligges et al., 2018) and *seewave* (Sueur et al., 2008) R packages, and RMS normalisation and MP3 exports were executed with Audacity.[3]

Each excerpt from the chills source was also tagged with a genre, with its associated excerpts from the matched source inheriting that tag. Note that this procedure still circumvented issues of genre labelling when determining stylistic preference, since these genre tags were only used to group excerpts into broadly similar categories in order to more quickly hone in on an ideal set of stimuli for each participant, as described in the next paragraph. Excerpts from the chills source were tagged with genre labels taken from MG-CT, the Music Genre-Clips Test (Bonneville-Roussy et al., 2017), by manually comparing the excerpts from the present study to the excerpts provided by the MG-CT and selecting the most closely related genre, resulting in excerpts being categorised into 11 different genres, ranging from two excerpts for punk music to 28 excerpts for classical music.

The online questionnaire for the stimulus allocation session was developed using *psychTestR* (P. M. C. Harrison, 2020), an R package allowing for the design

[3]https://www.audacityteam.org

of online experiments recruiting a more complex internal logic, using R code snippets, than that provided by most online survey platforms. The questionnaire was hosted on shinyapps.io,[4] and participant data was automatically uploaded to Dropbox[5] upon completion. Using psychTestR was necessary to dynamically adapt the order of the presented excerpts in order to select a set of 20 appropriate stimuli as fast as possible, since it was not reasonable to expect participants to listen to and provide ratings for all 372 available excerpts before making a selection. When taking the questionnaire, participants were first asked for their age and gender, before taking the Musical Training sub-scale of the Gold-MSI (Müllensiefen et al., 2014). They were then presented with the stimulus allocation task. For this task, participants were presented with a series of excerpts and posed two questions for each excerpt, asking them to report whether or not they knew the piece of music the excerpt was taken from (with a possibility to answer that they were unsure), and to rate how much they tended to like music which sounded like the excerpt on a five-point Likert scale ranging from "Dislike very much" to "Like very much". In the instructions before the test, participants were explicitly instructed that the second question referred to the genre or style of the excerpt, and that they were not asked whether or not they liked the excerpt itself.

The goal of the task was to select five unfamiliar excerpts for each combination of source (chills or matched) and stylistic preference (liked or disliked), while maximising the number of extreme values for stylistic preference (i.e., "Dislike very much" or "Like very much"). In order to do so, the order of presentation of the excerpts was determined using the following logic. First, two excerpts were sampled from the chills source for each tagged genre, and presented in a random order. If excerpts were familiar or familiarity was unsure, they were discarded. If excerpts were unfamiliar and rated with extreme values of stylistic preference, they were retained. Stylistic preference of the presented excerpts was averaged by genre and updated with each answer, allowing the questionnaire to identify the most liked and disliked genres for a given participant at any time. Then, more excerpts from the chills source were played, focusing on the most liked or disliked genres depending on which category had the fewest retained excerpts, in order to maintain the balance between the number of excerpts in very liked and disliked genres. Once 40 excerpts were played, ten excerpts were retained for both liked and disliked genres, and were locked in for the chills set, even if they gathered less extreme ratings, e.g., "Dislike somewhat" instead of "Dislike very much". Then, the task randomly iterated through the excerpts from the matched source that were paired with each retained excerpt, to try to

[4]https://www.shinyapps.io
[5]https://www.dropbox.com

identify similarly liked or disliked matches, starting with the most highly ranked match before trying the other two matches. Similarly, excerpts from the matched source were discarded if familiar or if familiarity was unsure, and were retained if they were unfamiliar and had extreme ratings of stylistic preference in the same direction as their associated excerpt from the chills source. If enough liked or disliked excerpts were retained, the task stopped trying to fill that category of stylistic preference. The task ended successfully if each combination of source and stylistic preference received five excerpts with extreme ratings of stylistic preference, or if 80 excerpts were played and excerpts with non-neutral ratings of stylistic preference could be allocated to each condition. Participants failed the task if it ran out of excerpts from the matched source, or if 80 excerpts were played and five excerpts could not be allocated to each condition.

Goosecam

Before describing the software used for the lab sessions, it is useful at this point to discuss the device that was used to record piloerection during these sessions. The device consisted of a slightly more compact Goosecam (see Chapter 2 for a brief description), intended to minimise the discomfort of wearing the rather bulky original version of the Goosecam, made of a $160 \times 40 \times 40$ mm aluminium bar with a cutout to accommodate a webcam (Benedek et al., 2010). The thorough specifications detailed by Benedek et al. (2010) allowed for the design of a smaller device which complied with the original requirements of filming the skin through a 40×40 mm cutout, illuminated from an angle of approximately 15º, from a distance of 46 mm. The body of the device was made of a laser-cut box using a 3 mm acrylic sheet, the inside of which was lined with matte black self-adhesive vinyl to minimise reflections. The pattern for that box was adapted from an open-source Raspberry Pi case pattern,[6] and is shown in Figure 3.1.

Inside the Goosecam, various components were powered and controlled by an Arduino Nano board: a white LED panel affixed to one of the sides of the body to provide directional light at an angle from the skin, a small OLED display to facilitate stimulus identification when processing the videos for the analysis, and a piezo buzzer and a green LED light to enable precise synchronisation between the video and audio signals. All of these components were kept in place with a healthy amount of hot glue to ensure they would not get in between the camera and the skin cutout. The camera for this version of the Goosecam was a GoPro Hero5 Session—a compact camera which allows native linear correction for the fisheye distortion that is common in many GoPro cameras. Finally, the edges of the bottom side of the box were padded with foam sheet to minimise

[6]https://github.com/diy-electronics/raspberrypi-b-plus-case

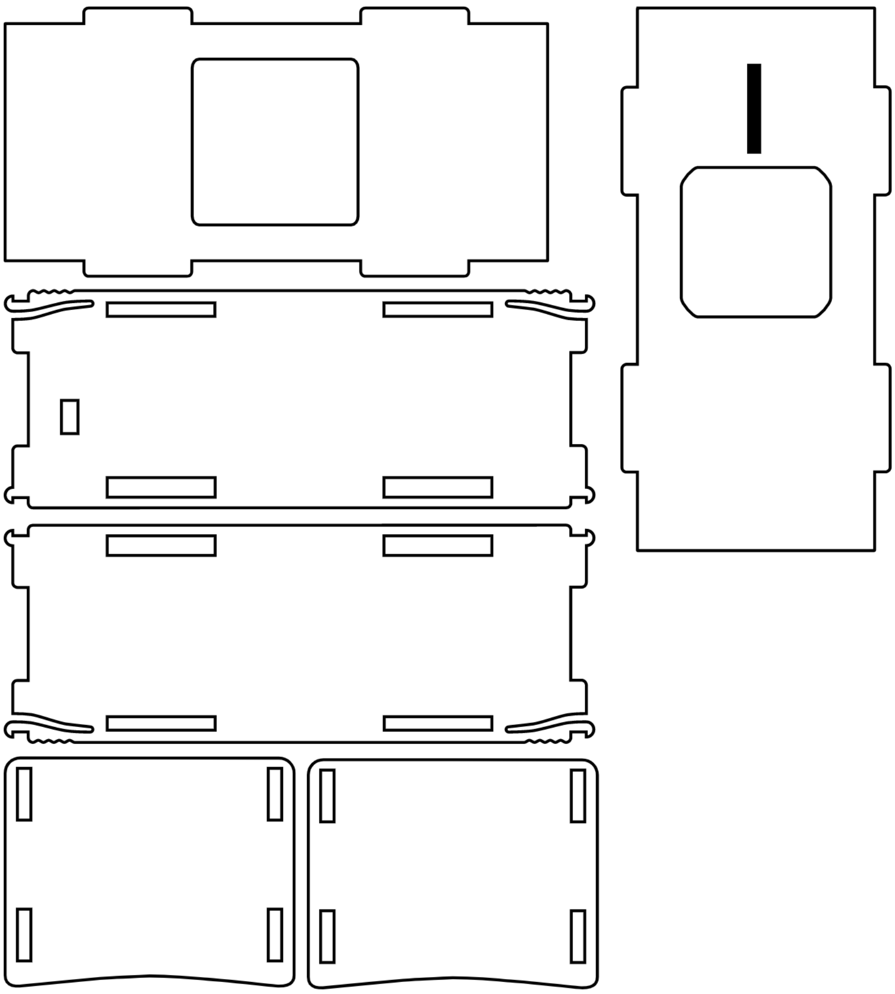

Figure 3.1: Laser-cut vector design for the compact Goosecam. The top panel shows the bottom side of the device, with a cutout for the skin to show through. The two panels below that show the long sides of the device, with pressure-activated locking mechanisms, cutouts for the top and bottom panels, and a cutout for the cable connecting to the Arduino board. The two bottom panels show the short sides of the device, with cutouts for the locking mechanisms. Finally, the panel on the right shows the top side of the device, with a cutout for the GoPro camera, and a groove in black for the camera mount.

participant discomfort, and the device was held in place by elastic velcro straps looped through the bottom cutouts of the side panels (see Figure 3.2).

An Arduino script controlled the board. It read data sent from the computer it was connected to via the board's serial port, and extracted from this data information about whether a track was about to start playing or had just stopped playing, as well as unique identifiers for participants, lab sessions, and tracks.

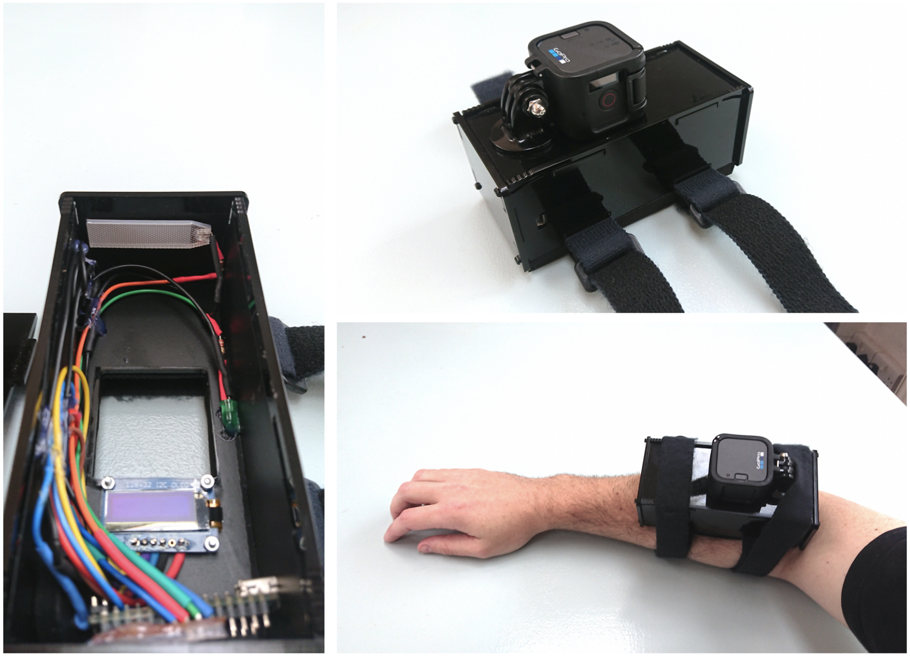

Figure 3.2: Pictures of the compact Goosecam. The picture on the left shows the internal components with, going top-down, the white LED panel, the green LED, the OLED display, the piezo buzzer hidden in the bottom-left corner, and the Arduino board. The other two pictures show the Goosecam with the GoPro mounted on top.

When a track started, the board instructed all indicators to switch on (white LED panel, LCD screen displaying all identifiers, green LED, and piezo buzzer). After one second, the components switched off except for the white LED panel which provided illumination for the skin. The panel switched off when a new serial message was received to indicate that a track had just ended.

Lab session platform

The objective of the first and last lab sessions was to collect continuous data as participants listened to full tracks. We were interested in gathering piloerection data, acquired via the Goosecam, as well as continuous self-reports for the occurrence of MECs and pleasure. While it would have been interesting to record self-reports of MECs intensity (using a slider, for instance) in conjunction with piloerection data, in order to evaluate whether or not piloerection occurs beyond a specific threshold of MEC intensity (see Chapter 2), it was decided to use button presses to record binary responses instead, for the purpose of the planned analyses and to minimise participant distraction given the complexity of the task.

The full tracks were converted to Ogg format with a 44.1 kHz sample rate and 16-bit bit depth, before RMS normalisation was applied using Audacity, with a target RMS level of -18 dB and linked stereo processing, ensuring that the original left/right stereo balance was maintained after normalisation. These steps were necessary to ensure consistent sound quality and levels when music was played at the same volume throughout the experiment (Mori & Iwanaga, 2014a).

The platform for the lab sessions was designed using OpenSesame (Mathôt et al., 2012), an open-source graphical experiment builder enabling the sequencing of graphical tasks based on code snippets written in Python. This platform was selected for its ability to record the precise timing of key strokes, and for the flexibility in its underlying logic and user interface design. The platform first presented participants with instructions about how to operate it, followed by a headphones volume calibration task, and a quick practice task (see Figure 3.3 for some of the instructions which were presented to the participants). The four actions participants were asked to take when listening to each track were to maintain the C key pressed down during experiences of MECs if any occurred, to maintain the M key pressed down whenever they found a musical passage pleasurable if at all, and after each track, to indicate whether or not they already knew the track they just listened to, and to rate how much they liked said track. The question about track familiarity was only presented in the first lab session, to confirm that the tracks were indeed unfamiliar following the online stimulus allocation task, and to present alternative tracks if not.

For each track, the following series of actions and logging events were taken by the platform. First, a serial message was sent to the Goosecam one second before recording a 30-second baseline measurement (required for processing Goosecam data), concluded with another serial message to signify that the baseline recording was over. The platform then slept for three seconds, before reiterating the procedure and playing the track instead of performing a baseline recording (see Figure 3.4 to see the platform during baseline recordings and when playing a track). While the track was playing, the onset and offset of key strokes were precisely recorded, in order to log when and for how long participants reported occurrences of MECs or pleasure in music. The accuracy and synchronisation of serial messages, track-playing triggers, and key stroke logging actions were extensively tested and validated prior to running the experiment. Note that the potential effect of key presses on the occurrence of piloerection was not tested, due to extensive previous evidence that such an effect is most likely not present (see Chapter 2).

Similarly to the stimulus allocation task, the order of presentation of the tracks followed some underlying logic, aimed at ensuring that three unfamiliar

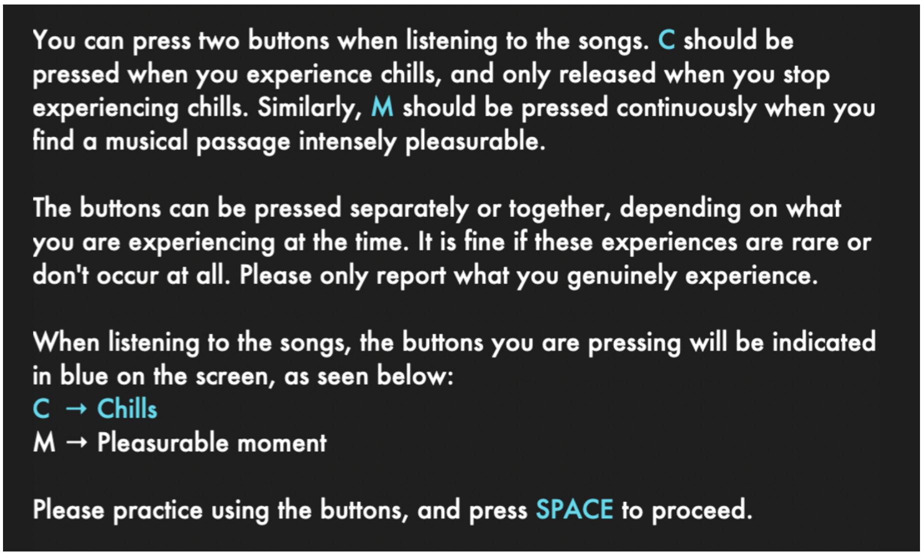

Figure 3.3: Graphical user interface for the lab session platform. This step of the participant instructions describes how to use key presses to report occurrences of MECs and pleasure in music throughout the experiment. In this case, the C key is pressed down, and is therefore highlighted in blue in the user interface.

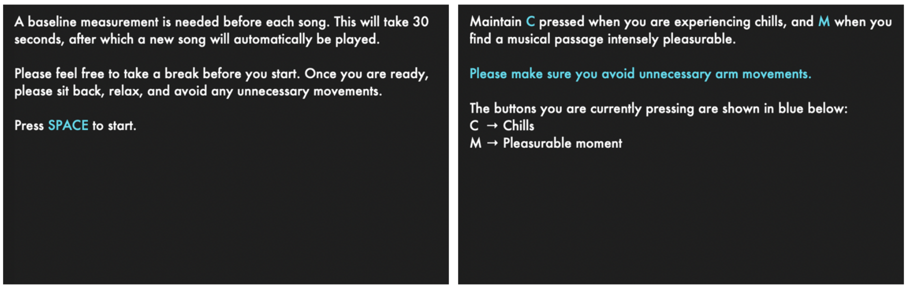

Figure 3.4: Graphical user interface for the lab session platform during the track listening task. Before each track, a baseline recording is performed. During each track, participants can use key presses to report occurrences of MECs and pleasurable moments in music. During both phases, participants are instructed to avoid unnecessary moments.

tracks would be presented for each combination of source and stylistic preference, which is why five tracks per condition were retained in the stimulus allocation task to allow room for some of the retained track actually being familiar to the participants. In practice, for both high and low stylistic preference, the platform randomly selected three tracks from the chills source and their associated tracks from the matched source, and played them in random order. If a track was identified as familiar, it was discarded along with its paired track, and that pair was replaced by one of the two back-up options from the stimulus allocation task. The task ended successfully when the participant had listened to three unfamiliar tracks in each condition, or failed when no more tracks were available.

3.2.3 Procedure

Participants were recruited through internal mailing lists of undergraduate and postgraduate students at Queen Mary University of London, and by emailing the participants from the survey study who accepted to be contacted for future research. In order to be eligible for taking part in the study, participants needed to often experience MECs (defined here as shivers, goosebumps, or a tingling sensation) when listening to music, to listen to at least an hour of music per day, and to have the ability to access Spotify on a computer or smartphone. 33 participants (18 female, 15 male), ranging in age from 19 to 38 years (M = 25.5 years, SD = 5.5 years), took the stimulus selection online task, during which they completed the Musical Training sub-scale of the Gold-MSI, and underwent the stimulus selection process, by rating up to 80 randomly selected, 15-second excerpts for familiarity and stylistic preference, resulting in an individualised set of 20 candidate tracks for each participant, comprising of five tracks for each combination of source (chills or matched) and stylistic preference (liked or disliked). The researchers were automatically notified of test completion, and followed up with the participants to schedule a time for the first lab session.

Of those 33 participants, 3 were discarded because a set of stimuli could not be finalised for them, either because too many excerpts were rated as familiar, or too few excerpts were rated with non-neutral values for stylistic preference. As a result, 30 participants (16 female, 14 male), ranging in age from 19 to 36 years (M = 24.9 years, SD = 4.8 years), took part in the first lab session, conducted in a soundproofed listening room a few days after the online task. They were asked to adjust headphones volume to a comfortable level, before being allowed to experiment how to interact with the platform. They were then presented with 12 to 20 retained tracks from the online task, while using keyboard presses to continuously report MECs and pleasure in music, if any occurred. Piloerection was recorded using the Goosecam, worn on each participant's non-dominant arm, and participants were therefore asked to avoid unnecessary arm movements when listening to the tracks. They were given opportunities to take a break between each track. At the end of the lab session, participants were offered the option to take part in the longitudinal phase of the study.

Of those 30 participants, 13 (7 female, 6 male), ranging in age from 19 to 36 years (M = 25.6 years, SD = 5.1 years), agreed to take part. They were emailed a link to a Spotify playlist containing the 12 unfamiliar tracks they listened to during the lab session, as well as a link to a short, password-protected, Qualtrics survey. They were asked to listen to the playlist eight times (but not more than once a day), away from the lab, before the final lab session, which was scheduled two weeks after the first lab session. They were instructed to avoid distracting social contexts when listening to the playlist, whenever possible, but were also told that intense, focused listening was not necessary (e.g., listening when commuting, exercising, etc. was accepted), and that it was fine to not listen to the entire playlist at once, as long as it was listened to in full before completing a survey. Participants completed a survey after each time they finished the playlist, in order to report the tracks during which they experienced MECs, if any. The tracklist order was manually randomised by the researcher every five days, and participants were also given the option to receive email reminders to complete the tasks at regular intervals. The duration of the longitudinal phase and number of repetitions were chosen to be broadly consistent with other longitudinal studies investigating psychological responses to music (Chmiel & Schubert, 2019; Grewe et al., 2007; Madison & Schiölde, 2017), while trying to minimise attrition and fatigue.

The final lab session was essentially similar to the first session, with the exception that participants were not asked to rate familiarity for each track after listening to them.

Participants who completed both the stimulus selection task and the first lab session were entered into a draw for a £150 Amazon voucher. In addition, participants who completed the full study were offered £20 for their time. It is worth noting that the sample sizes for the different parts of the present study were fairly small, which is due to the time-consuming nature of the experiment and the monetary constraints for participant compensation. We would note, however, that these sample sizes are consistent with similar previous research on piloerection or longitudinal effects (e.g., Bannister & Eerola, 2018; Madison & Schiölde, 2017; Wassiliwizky, Jacobsen, et al., 2017).

A summary of all the data collection steps is shown in Table 3.1 and a graphical representation of the experimental design is shown in Figure 3.5. The full experiment was tested on three volunteers. This resulted in the identification and correction of small software bugs for some edge cases, in validating the Goosecam by manually confirming the consistency between video input and device output, and more importantly, in switching from using a single Goosecam baseline for the whole experiment to recording a baseline before each track. This change resulted in more consistent measurements, and had the added advantage of allowing participants to take a break during the experiment without disrupting the calibration of the Goosecam.

Table 3.1: Summary of data collection steps

Step	Data collection	Participants	Objective
1. Online survey*	Feb 2018 – May 2018	221 online (Qualtrics)	Compile a list of tracks which elicit MECs
2. Stimulus allocation	Nov 2018 – Jun 2019	33 online (psychTestR)	Select unfamiliar tracks from *Step 1* for each participant
3. First lab session	A few days after *Step 2*	30 in-person from *Step 2*	Collect continuous data as participants listened to full tracks
4. Longitudinal phase	Immediately after *Step 3*	13 from *Step 3*	Increase familiarity with previously unfamiliar tracks
5. Final lab session	Two weeks after *Step 3*	Same as *Step 4*	Assess the effect of familiarity on MECs

Note. * The online survey was left running after May 2018 to build a dataset for the computational study detailed in Chapter 5.

3.3 Analysis

3.3.1 Piloerection data

Video file processing was automated using FFmpeg (Tomar, 2006), an open-source command-line tool for handling multimedia files. Due to the long duration of the lab session, the footage was split into multiple files when saved to the GoPro's SD card. Using FFmpeg, the files were concatenated into a single, large video file per session. Synchronisation between the video and the experiment logs was achieved by careful manual identification of the timestamp at which the Goosecam's green LED light and piezo buzzer first switched on, signifying that the serial message for the first baseline recording had just been received, and by trimming the video before that timestamp. The skin-side cutout was

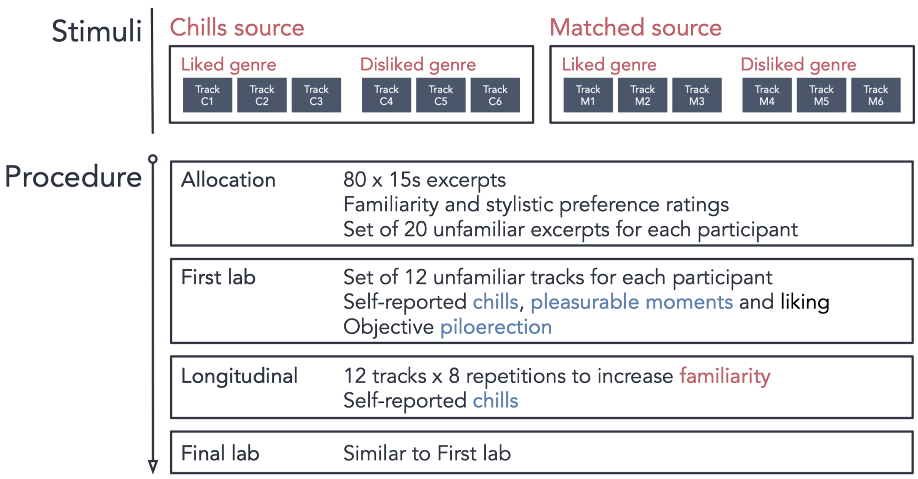

Figure 3.5: Simplified summary of the experimental design. The main independent variables are shown in red, and the main dependent variables are shown in blue.

centred and cropped down to 506 $\times$ 506 px to ensure only an illuminated portion of the skin would be captured. The video was then rotated by 90º to reach the expected orientation (due to the way the GoPro had to be mounted on the Goosecam), audio signal was removed from the file, and the frame rate was downsampled to 10 Hz. Finally, the video was automatically split into individual files corresponding to each track, using the timestamps logged by the lab session platform.

Individual video files were processed with Gooselab,[7] a MATLAB toolbox which performs the computations described in Benedek et al's (2010) article. Broadly speaking, for each frame, the toolbox extracted the largest possible square image from the video file (an unnecessary step in this case, since the frames had already been cropped), before converting it to grey scale, applying a high-pass filter, running a two-dimensional discrete Fourier transform, performing angular averaging, and computing piloerection as the maximum amplitude within a specific frequency range. This process resulted in a time-series showing piloerection values for each frame, with higher values corresponding to greater observed piloerection.

Piloerection events then needed to be extracted from these continuous time-series. To do so, for each combination of lab session, participant, and track, a threshold was set at three standard deviations over the average baseline value. Piloerection events for each track were then assigned to each period of at least ten consecutive frames (i.e., one second) exceeding that threshold value, with the onset and offset of piloerection occurring on the first and last frame the threshold was exceeded for that event, following the updated recommendations by Benedek and Kaernbach (2011). Figure 3.6 shows a particularly clear episode of piloerection, overlapping with self-reported MECs and pleasure.

3.3.2 Statistical analyses

After data processing, the three dependent variables (MECs, piloerection, pleasure) consisted of time-series of binary events. There were several possibilities with regard to quantifying their occurrence within each track. For the first set of analyses, we opted to sum the duration of each event for each track, and to scale that value according to track duration, so as to not penalise slightly shorter tracks. The resulting value can be interpreted as the total duration (in seconds) for which events (e.g., MECs) were experienced during a track, if that track were five minutes long. This approach was chosen instead of simply counting tracks during which events occurred or not, since most tracks featured reports of MECs

[7] The official link to the Gooselab toolbox is deprecated, and currently, there seems to be no official repository for the toolbox but luckily, a working version was found on Github: https://github.com/tstenner/gooselab

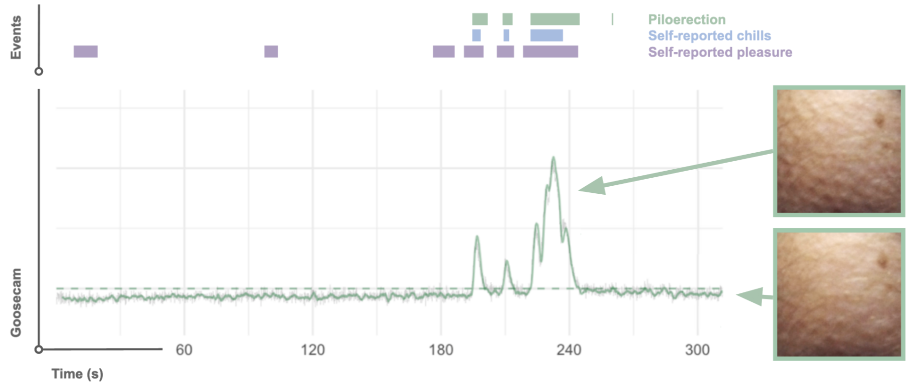

Figure 3.6: Time-series of piloerection values. The threshold for a piloerection event to occur is indicated by the green dashed line. Raw piloerection values are shown faintly in grey, and smoothed values in green, for visualisation purposes only. If piloerection values exceeded the threshold for ten consecutive frames, a piloerection event was assigned to the track, as indicated by the green bar. In this case, the piloerection events considerably overlapped with self-reported MECs, in blue, and pleasure, in purple. Example video frames are shown at two different time points, displaying piloerection in the top frame, and no piloerection in the bottom frame.

and pleasure, as later explained in the results, which would have resulted in a loss of much of the signal in the data. Another option would have been to simply count the number of events per track, but this would have been biased by small imprecisions in the measurements, such as long events being interrupted by the self-report key not being pressed properly for a short moment, or by piloerection values dipping below the threshold for just a few frames, for instance. Summing up the duration of events seemed like a reasonable compromise to keep as much of the signal as possible while minimising the risk for bias. Note that for these analyses, the resulting variable for piloerection was log-transformed in order to adhere to the assumptions of the statistical tests that were used.

For this first set of analysis, we fit a linear mixed-effects model for each of the dependent variables, using familiarity, stylistic preference, and source as fixed effects, and participant and track as random effects. We opted not to test for interaction effects due to the relatively large number of variables considering the small sample size, which would have increased the risk of overfitting the models. Since these models included all participants, regardless of whether or not they completed the longitudinal phase of the study, we also conducted paired t-test to test the effect of familiarity on each dependent variable for the subset of participants who attended both lab sessions. Finally, we fit another linear mixed-effect model to assess the extent to which the combination of all variables predicted the variance in the ratings of liking collected after listening to each track, therefore using MECs, piloerection, pleasure (all preprocessed by scaling the features, since they were continuous), familiarity, stylistic preference, and source as fixed effects, and participant and track as random effects. Note that for this analysis, the emphasis was placed on predictive power rather than interpretability, seeing as some of the variables were highly collinear.

All mixed-effects were fit using the *lme4* R package (Bates et al., 2015). Diagnostic tests revealed that there were no high-leverage outliers, and that there was adequate homoscedasticity and normality of residuals (except for piloerection, at first, which is why the variable was eventually log-transformed). Model fits were evaluated by performing likelihood ratio tests, which were run by conducting ANOVAs comparing the fitted models with null models including random effects only. Marginal and conditional R^2 values were obtained with the *MuMIn* R package (Bartoń, 2020). Marginal R^2 outlines the variance explained by the fixed effects only, while conditional R^2 outlines the variance explained by the whole model (Nakagawa & Schielzeth, 2013). Finally, p-values for the fixed effects were obtained using the *lmerTest* R package (Kuznetsova et al., 2017).

To conclude the first set of analyses, a range of tests were conducted as sanity checks, as well as to explore some relationships in the data. First, we ran correlation tests to verify whether or not room temperature had an effect of the occurrence of piloerection and self-reports of MECs. Second, we ran a series of correlation tests and t-tests to test the existence of relationships between age, gender, and musical training and all of the dependent variables. Finally, we tested the correlations between the dependent variables themselves.

The second set of analyses concerned the longitudinal phase of the study, for which only occurrence of MECs at the track level was collected. Note that this data was analysed separately due to the completely different ways in which data was collected. While the longitudinal phase was mainly aimed at increasing familiarity for the final lab session, it also had the potential to reveal how repetition affected the occurrence of MECs. For this analysis, we fit a logistic mixed-effect model with number of repetitions as a fixed effect and participant, source, and stylistic preference as random effects. Apart from diagnostic tests on residuals, which are much less interpretable for logistic regression, the same procedures as detailed above were used to explain the model.

For the final set of analyses, the dependent variables were taken in their binary time-series form. The purpose of these analyses was to assess the degree of overlap between MECs, piloerection, and pleasure. For this purpose, permutation tests were chosen, as they allow for the convenience of choosing the most appropriate test statistic for the task at hand, and their nonparametric nature eliminates the need to validate many assumptions inherent to time-series analysis. A one-sided permutation test was run for each pair of dependent variables, using Monte Carlo estimation with 5000 replications. The test statistic was calculated by taking the percentage of frames during which both events occurred simultaneously, for each track, and then averaging the obtained values for each track over the whole experiment, resulting in an average rate of overlap per track. Permutations were conducted by randomly rearranging event onsets within each track, while ensuring that events did not overlap within a single variable. For example, if a given track was 20 seconds long, and contained two reports of MECs, from 0:05 to 0:07 and 0:11 to 0:16, a valid permutation for this variable would be to rearrange MECs so that they occur from 0:01 to 0:06 and from 0:17 to 0:19. However, rearranging MECs so that they occur from 0:01 to 0:06 and from 0:05 to 0:07 would not be valid, since these events overlap with each other.

3.4 Results

One participant displayed unusual patterns in their self-reported responses. After close inspection of their data, it appeared that the participant did not fully understand the task, and repeatedly pressed keys when MECs or pleasure occurred instead of keeping the keys pressed down as instructed, resulting not only in many more key strokes than any other participant, but also in small bugs due to the speed at which data was being written by OpenSesame while running the task. Instead of manually fixing the data, which would have required making a lot of assumptions as to what was the intended behaviour, we deemed it more appropriate to exclude this participant from the data analysis.

Of the remaining 29 participants, 27 reported MECs at least once, 11 experienced piloerection at least once, and all of them reported pleasure at least once. When looking at data at the track level, out of the 152 tracks that were listened to throughout the experiment, 116 caused MECs at least once, 32 for piloerection, and 135 for pleasure. Overall, using the scaled measure for the dependent variables (i.e., total duration per track if each track were five minutes long), MECs were experienced 20.8 seconds per track on average, 3.10 seconds for piloerection, and 46.0 seconds for pleasure.

Effects of source, stylistic preference, and familiarity

For MECs, the linear mixed-effects model yielded a significant fit ($\chi^2(3) = 36.84$, $p < .001$, $R^2_m = .047$, $R^2_c = .502$), revealing significant effects on MECs for familiarity ($b = 8.23$, $p = .013$), stylistic preference ($b = 15.47$, $p < .001$), but not source ($b = 4.80$, $p = .123$), with average reports of MECs increasing by 9.6 seconds for familiar tracks, and by 15.4 seconds for tracks in a liked genre.

For log-transformed piloerection, the linear mixed-effect model yielded a significant fit ($\chi^2(3) = 11.35$, $p = .010$, $R^2_m = .017$, $R^2_c = .436$), revealing significant effects on piloerection for familiarity ($b = -0.19$, $p = .013$), stylistic preference ($b = 0.14$, $p = .031$), but not source ($b = -0.06$, $p = .394$), with average detected piloerection decreasing by 4.0 seconds for familiar tracks, and increasing by 2.75 seconds for tracks in a liked genre.

For pleasure, the linear mixed-effects model yielded a significant fit ($\chi^2(3) = 65.43$, $p < .001$, $R^2_m = .088$, $R^2_c = .527$), revealing a significant effects on pleasure for stylistic preference ($b = 37.42$, $p < .001$), but not familiarity ($b = -0.43$, $p = .931$) or source ($b = 2.46$, $p = .599$), with average reports of pleasure increasing by 39.3 seconds for tracks in a liked genre.

Considering the large differences between marginal and conditional R^2 in each model, simple linear regression models were also run to assess the individual contributions of participant ID and track ID to the variance in the dependent variables, revealing that, on their own, participants explain 45% of the variance in MECs, 30% in piloerection, and 38% in pleasure, and tracks explain 10%, 18%, and 21% respectively, as measured by the adjusted R^2 of the fitted models.

Restricting the analysis only to the participants who completed both lab sessions, paired t-tests revealed a significant effect of familiarity (U for *unfamiliar*, F for *familiar* in the reported summary statistics) on MECs ($t(148) = 2.14$, $p = .034$, $M_U = 20.31$, $SD_U = 42.34$, $M_F = 28.50$, $SD_F = 52.46$) and piloerection ($t(148) = -2.11$, $p = .037$, $M_U = 4.38$, $SD_U = 23.03$, $M_F = 0.40$, $SD_F = 2.96$), but not on pleasure ($t(148) = -0.08$, $p = .937$, $M_U = 44.63$, $SD_U = 62.15$, $M_F = 44.29$, $SD_F = 57.52$), confirming the results from the mixed-effects models. For those participants, familiarity increased average reports of MECs by 8.2 seconds, and decreased detected piloerection by 4.0 seconds, consistently with the previous analyses.

Combined effects on liking

For liking, the linear mixed-effects model yielded a significant fit ($\chi^2(6) = 328.42$, $p < .001$, $R^2_m = .505$, $R^2_c = .709$), revealing significant effects on the ratings of liking collected after listening to each track for continuously reported pleasure ($b = 0.91$, $p < .001$) and stylistic preference ($b = 1.90$, $p < .001$), but not for the other variables, although it is worth keeping in mind that these individual effects are not interpretable, due to the high collinearity between some of the predictors.

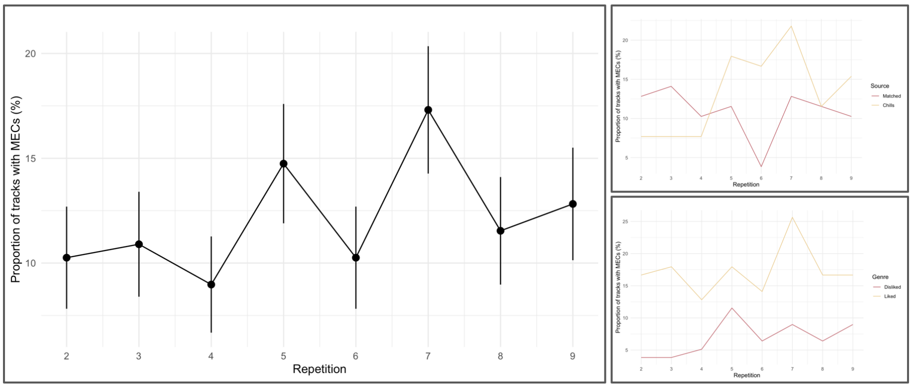

Figure 3.7: Effect of repeated listening on the occurrence of self-reported MECs during the longitudinal phase of the study, starting from the second to the ninth repetition (the first and last repetitions occurring during the lab sessions, and therefore not shown or analysed). The first plot shows what might look like a slight overall increase over time, but with large day-to-day swings and high variance, as shown by the error bars. The two smaller plots show the differences in occurrence of MECs when tracks are split by source or stylistic preference, demonstrating that no clear patterns emerged other than that of tracks in liked genres causing more MECs than in disliked genres overall.

Sanity checks

Room temperature was 21.2ºC on average (SD = 1.9ºC), which is comparable to other studies which reported temperature (Benedek & Kaernbach, 2011; Laeng et al., 2016), and was not correlated with either MECs ($r(35) = -0.06$, $p = .711$) or piloerection ($r(35) = -0.09$, $p = .599$).

Age, gender, and musical training had no effect on any of MECs, piloerection, and pleasure (all $p > .05$), with gender (F for *female*, M for *male* in the reported summary statistics) trending the closest to significance, as seen with t-tests on MECs ($t(18.16) = 1.87$, $p = .077$, $M_F = 9.98$, $SD_F = 18.15$, $M_M = 29.65$, $SD_M = 33.59$), piloerection ($t(13.68) = -1.48$, $p = .163$, $M_F = 6.23$, $SD_F = 13.56$, $M_M = 0.81$, $SD_M = 2.12$), and pleasure ($t(20.03) = 1.88$, $p = .074$, $M_F = 34.06$, $SD_F = 30.76$, $M_M = 63.72$, $SD_M = 48.61$).

Effect of repeated listening

For the effect of repeated listening on MECs, the logistic mixed-effects model did not reach a significant fit ($\chi^2(1) = 1.92$, $p = .166$, $R^2_{m\Delta} = .001$, $R^2_{c\Delta} = .263$), and there was no significant effect ($b = 0.06$, $p = .16$), presumably due to high variance resulting from low sample size, as seen in Figure 3.7.

Relationships between dependent variables

Correlation tests showed that average event durations were not correlated between piloerection and either of MECs ($r(491) = -0.02$, $p = .658$) or pleasure ($r(491) = 0.01$, $p = .827$), but that they were correlated between MECs and pleasure ($r(491) = 0.52$, $p < .001$).

When taken as binary time-series, however, these three variables showed highly significant overlaps, as shown by one-sided permutation tests using Monte Carlo estimation with 5000 replications. On average, the number of frames showing overlap between both responses represented 0.11% of the total frames in the track for MECs and piloerection ($p < .001$), 3.26% for MECs and pleasure ($p < .001$), and 0.16% for piloerection and pleasure ($p < .001$). It should be noted that these effects seem small in magnitude because overlapping frames were compared to the total number of frames in all tracks. The permutation tests revealed that this degree of overlap was much greater than what would be expected by chance.

3.5 Discussion

In this study, we investigated the effects of musical content, stylistic preference, and familiarity on the occurrence of MECs, piloerection, and pleasure, by setting up a controlled, longitudinal experiment, ensuring that adequate comparisons could be drawn across all variables.

First, we observed that the majority of participants reported MECs and pleasure throughout the experiment, but than less than half of them experienced detectable piloerection. This represent an unusually high proportion of MECs (see Chapter 2 for a discussion about prevalence in experimental contexts, e.g., Colver & El-Alayli, 2016; Grewe et al., 2009a; Konečni et al., 2007), which could be explained by the fact that each phase of the study required a lot of music listening, and that the study itself was particularly long. In addition, participants were recruited based on their ability to often experience MECs, and tracks were selected to maximise the occurrence of MECs, or closely match tracks that do. Conversely, it could be argued that the duration and format of the experiment encouraged participants to overly report MECs. For instance, a participant could find it unusual to listen to twelve or more full tracks during an experiment while sitting in front of a keyboard and doing nothing at all, if they didn't experience MECs. This concern is slightly alleviated by the fact that, when they were reported, MECs did significantly overlap with piloerection, which is not consciously controlled by the participants.

The fact that participants experienced much fewer episodes of piloerection suggests a disconnect between reported MECs and piloerection. While there are arguments that piloerection might only occur beyond a specific MEC intensity threshold (Sumpf et al., 2015), other explanations could be that current piloerection detection methods are not yet accurate enough, or that piloerection and MECs are not mutually exclusive. It is difficult to determine from this study alone which of the responses is a more reliable indicator of MECs in general. More research is needed on the exact relationship between self-reported MECs and piloerection, and the methods introduced in this chapter might help in conducting such research. For the present study, however, this discrepancy between MECs and piloerection needs to be taken into consideration when interpreting the results discussed below.

Stylistic preference had a positive effect on all three of MECs, piloerection, and pleasure, which is not particularly surprising when it comes to pleasure, since we would expect more pleasure to be experienced when listening to music in liked genres. However, the fact that stylistic preference also drove the occurrence of MECs and piloerection suggests that the conflicting results in previous research (Bannister & Eerola, 2018; Nusbaum & Silvia, 2011) might have been due to inadequate research methods. In other words, rather than preference for specific genres (as identified by STOMP, for instance) leading to more MECs being experienced, it could be that it was the interaction between a listener's stylistic preferences and the music being listened to that had such an effect. This raises questions for future research about the role of stylistic knowledge or exposure. For instance, would MECs occur in liked but unfamiliar genres? The effects of stylistic preference and stylistic knowledge might be independent, or one could be mediated by the other, such as stylistic knowledge leading to stylistic preference, and in turn, stylistic preference increasing the likelihood of MECs, for instance. Further systematic study could help in this regard, and could potentially uncover valuable information about the role of expectation in the experience of MECs.

More surprisingly, familiarity had opposite effects on MECs and piloerection, and no effect on pleasure. These results were confirmed when only looking at the cohort of users who completed the longitudinal phase of the study. Pleasure remaining stable with repeated exposure goes against previous longitudinal findings that familiarity increases liking and music preference (Chmiel & Schubert, 2019; Madison & Schiölde, 2017). However, in the context of this study, pleasure was disproportionately higher for tracks in liked genres than for those in disliked genres. Pleasure could therefore have reached extreme values at the beginning of the study, and remained stable in time. We would also argue that pleasure is a more intense subjective response than liking or preference, and as such, might evolve more slowly over time. The opposite effects of familiarity on MECs and piloerection is interesting, and suggests, once again, fundamental differences between the two responses. One possible interpretation is that piloerection is more reliant on physiological processes and might be more subject to habituation as a result. Or it could be that the quantity of MECs increases with repeated exposure, but their intensity decreases, therefore not exceeding the threshold for piloerection discussed earlier. Leveraging the framework identified in Chapter 2, these results could also suggest that MECs which involve piloerection rely on different psychological and evolutionary mechanisms than those which don't. Some MECs based on personal meaning could therefore benefit from repeated exposure, while others based on surprise could temporarily disappear due to frequent listening.

Musical content, manipulated via stimulus source, had no effect whatsoever. As will be discussed more extensively in Chapter 4, it could be the result of the strict matching procedure, i.e., an effect might have been detected by randomly sampling control tracks from any artist and level of popularity instead. Since it is fundamentally impossible to know if tracks from the matched source can cause MECs or not (in all likelihood, some of them can), it could be that, by applying a strict matching procedure, the features which cause MECs were present to the same extent in the chills and matched sources. The assumption behind the manipulation of stimulus source was that, if an effect of musical content exists, it should be detectable in aggregate because some stimulus-driven properties would make tracks from the chills source slightly more likely to cause MECs for several individuals.

Additional context comes from the fact that the models discussed so far all featured high conditional R^2 values, and low marginal R^2 values. This suggests that, while the results presented so far were indeed significant, they accounted for a small proportion of the variance in the dependent variables. Conversely, on their own, participants and tracks accounted for much variance in the outcomes, which indicates that participants have predictable response patterns, and that specific tracks tend to elicit similar responses. The former might be explained by individual differences, or simply by differences in how participants understood the task or decided what passes the threshold of reporting MECs or pleasure. Alternatively, the large number of tracks with no recorded response at all might on its own account for much of these conditional R^2 values.

Regardless, in the context of this study, it appears that likelihood of detecting an effect of musical content was overestimated, especially if MECs partly rely on personal meaning and not on stimulus-based elicitors, which would lead to ceiling effects. However, we do know through previous research (e.g., Sloboda, 1991) that there is an effect of acoustic and musical elicitors. Considering the small effect sizes obtained in the present analyses, it becomes clear that this effect is subtle, and that much larger amounts of data would be needed to detect it. This is a type of problem that is particularly well suited to computational approaches, and that will be the main subject of the rest of this thesis.

As opposed to the models discussed so far for continuous self-reports of MECs, piloerection, and pleasure, effect sizes for both the fixed effects and the whole model itself were large with regard to static ratings of liking collected after listening to each track, with source, stylistic preference, familiarity, MECs, piloerection, and pleasure accounting for about 50% of the variance in self-reported liking for tracks. While further analysis, and probably a different study design, would be needed to fully disentangle the individual contributions of these fixed effects, it is interesting that much of music preference can be predicted from a relatively small number of factors, and certainly suggests that preference could be even more accurately predicted in a study dedicated to this purpose.

As expected (see Chapter 2), there were no effects of age, gender, or musical training on any of the dependent variables, but interestingly, there were some subtle differences due to gender, with female participants experiencing more piloerection but fewer MECs and less pleasure than male participants. Interpretation is limited here by the fact that these differences were non-significant, and that there is currently no convincing hypothesis as to why such differences would exist.

For repeated exposure, while the primary objective of making participants listen to the tracks between the two lab sessions was to increase familiarity, we did collect some data about the occurrence of MECs. No effect of repeated listening was detected, but a small upward trend was visible on the figure representing the proportion of MECs experienced over time. It is worth noting that for this phase of the study, sample size was relatively low, there was less control over experimental conditions, and MECs were reported in a different way. Such an investigation might reveal interesting insights if conducted as a larger study, but for the present study, the manipulation achieved the intended objective of increasing familiarity.

Finally, we observed correlations between MECs and pleasure, but neither of them were correlated with piloerection. However, permutation tests revealed significant overlaps for all three variables. In fact, the test statistics had the most extreme values out of all 5000 Monte Carlo simulations for all three pairs of variables, indicating that the degree of overlap between them was almost impossible to achieve by chance. This suggests that while piloerection differs from MECs and pleasure, it is still strongly related to both of these responses. In other words, they overlap significantly, but this doesn't mean that they overlap exclusively, which provides further support for the hypotheses that piloerection requires exceeding an activation threshold (Sumpf et al., 2015), or alternatively, that they only represent a subset of MECs and pleasurable moments in music. A potential caveat is that the overlap between MECs and pleasure could be driven by asking participants to self-report the two measures at the same time, using similar modalities. The mutual overlap with piloerection alleviates that concern to some extent (but not fully), since piloerection was recorded using a device the participants had no control over, as discussed earlier. For the present study, it was not deemed practical to collect self-reports of MECs and pleasure separately, because repeated presentation of the stimuli could itself have affected MECs or pleasure, and using separate participant groups was not realistic given the large contribution of individual factors to the experience of MECs.

Overall, there remain some unexplainable differences between MECs and piloerection, with many participants reporting MECs, but few experiencing piloerection. It could be that the Goosecam is at fault, but visual inspection of the video files did not result in the detection of any obvious sensor errors. When debriefing after the end of the experiment, several participants mentioned experiencing piloerection in the neck, as has already been identified in previous research (e.g., Panksepp, 1995). While it is difficult to imagine asking participants to wear the current iteration of the Goosecam around their neck, it would be helpful to eventually have access to better, more compact sensors, which could enable the simultaneous collection of piloerection data from different areas of the body. This is related to the difficulty in gaining confidence in participants' self-reports of MECs, given the large differences in reporting behaviours across participants. Perhaps subjective experiences of MECs are very different across individuals, but we suspect that instead, it is just particularly difficult to identify a definition of MECs that is not liable to subjective interpretation. Further efforts towards the development of accurate and objective detection methods for the overall experience of MECs, as opposed to piloerection only, would go a long way in furthering research on MECs.

This study achieved many of its stated objectives, identifying strong effects of stylistic preference on MECs, piloerection, and pleasure, opposite effects of familiarity on MECs and piloerection, high predictability of liking based on a small set of variables, and significant overlap between MECs, piloerection, and pleasure, highlighting the close links between MECs and emotional and aesthetic responses.

The investigated phenomena are complex, and differ highly across individuals, and yet, these effects were detected, despite understandably small effect sizes. The only exception is for musical content, which most likely requires much larger amounts of data. Approaches aiming to do so are detailed in the next two chapters of this thesis, starting in Chapter 4 with a corpus analysis of the relationship between valence and MECs. Additionally, the present chapter focused on exploring open issues presented in Chapter 2 as opposed to explicitly investigating the possible pathways included in the preliminary model of MECs. Chapters 4 and 5 more closely explore these pathways and their associated elicitors and mechanisms.

Chapter 4

Perceived valence

In the study presented in Chapter 3, we didn't manage to identify behavioural differences when comparing tracks which have been previously reported as being able to cause MECs with tracks which were matched to those by artist, duration, and popularity. We hypothesised that this might have been due to small effect sizes, not detectable in behavioural experiments with relatively small sample sizes, and that applying computational methods to larger amounts of data might have the potential to identify differences between these two sets of tracks.

This chapter presents the application of such methods to the investigation of the relationship between valence on MECs. More specifically, since valence is a readily available feature for computational analysis, as explained below, it was chosen as a proof of concept to justify further computational work, as presented in Chapter 5. With this study, we aimed to disentangle findings that MECs have been linked to both happiness and sadness expressed by music. We conducted a computational analysis on a corpus of 988 tracks previously reported to elicit MECs, by comparing them with a control set of tracks matched by artist, duration, and popularity, hereafter referred to as tracks from the matched source. We analysed track-level audio features obtained with the Spotify Web API[1] across the two sets of tracks, resulting in confirmatory findings that tracks which cause MECs were sadder than tracks from the matched source, and exploratory findings that they were also slower, less intense, and more instrumental than tracks from the matched source on average. We also found that the audio characteristics of tracks from the chills source were related to the direction and magnitude of the difference in valence between the two sets of tracks. We discuss these results in light of the current literature on valence and MECs in music, provide a new interpretation in terms of personality correlates of musical preference, and review the advantages and limitations of our computational

[1] https://developer.spotify.com/documentation/web-api

approach.

4.1 Introduction

As discussed in Chapter 2, particular attention has been given to elicitors of MECs, resulting in the identification of a range of acoustic and musical features usually associated with the experience of MECs, such as sudden dynamic changes, increased roughness, crescendi, or the entrance of new instruments. These features represent local musical and auditory events, and are therefore reflective of the fleeting nature of MECs.

In contrast to continuous changes within musical stimuli, emotional characteristics of entire musical pieces have also been investigated. As a result, MECs have been associated with perceived valence. However, there is disagreement about the direction of this relationship. While Grewe et al. (2011) reported an increase in frequency of MECs for positively valenced music, Panksepp (1995) found that MECs were more frequently associated with perceived sadness. In the latter study, however, both happy and sad music were found to elicit MECs, reflecting subsequent findings that MECs are associated with both emotions when they are expressed by music (Bannister, 2020b; Mori & Iwanaga, 2017; Panksepp, 1995).

Conflicting effects of valence on MECs have been discussed in the context of being moved, a mixed emotional state involving sadness and joy (Menninghaus et al., 2015). More specifically, being moved has been associated with MECs when listening to music (Bannister, 2019, 2020b; Bannister & Eerola, 2018; Benedek & Kaernbach, 2011), and has been found to mediate the relationship between liking and sadness in response to music (Vuoskoski & Eerola, 2017). Moving stimuli often feature narrative displays of social separation or reunion (Wassiliwizky et al., 2015), prosocial behaviour (Wassiliwizky, Jacobsen, et al., 2017), or self-sacrifice (Konečni et al., 2007), but it remains unclear how such narrative features translate to music, and how stimulus valence relates to the occurrence of MECs. It could be that sad music provides an emotional context more conducive to the occurrence of MECs (Panksepp, 1995). Another plausible explanation for the mixed effects of valence on MECs comes from the possibility that MECs encompass several phenomenologically distinct experiences, partly characterised by different degrees of felt emotions (Bannister, 2019; Maruskin et al., 2012).

This perspective was further developed in Chapter 2, in which a preliminary model suggests three different pathways for the experience of MECs, linking different types of elicitors to the combination of psychological and evolutionary mechanisms most likely to elicit MECs, if not different types of MECs. In one such pathway, individuals with high trait empathy are suggested to be more receptive to emotional elicitors of MECs, such as perceived valence, leading them to mimic the perceived emotion through a process called emotional contagion, and then to experience MECs through the process of being moved. Other pathways link acoustic and musical elicitors to processes involving arousal and musical expectation, respectively. There is little research on the existence of these pathways, and confirming or refuting an effect of perceived valence on MECs would be a useful step in providing support for the existence of one of them, leading to a better understanding of the causes of MECs, and in turn, of music appreciation in general.

In light of the current evidence, it remains difficult to establish the role of expressed stimulus valence on the incidence of MECs. While behavioural approaches have contributed to identifying conflicting effects of happiness and sadness, they remain limited due to the number of stimuli which can be reasonably presented to participants, ranging here from 3 (Bannister & Eerola, 2018) to 23 (Grewe et al., 2011). Computational methods, however, can overcome such restrictions, and are well suited to the study of a large collection of naturalistic stimuli, at the cost of reduced control over experimental conditions. When we ran this study, there had been, to our knowledge, no use of corpus-based analysis in research on MECs, despite the success of similar approaches in research on music and emotion (e.g., Eerola, 2011). The analysis presented in this chapter was an attempt at addressing this gap in the literature, focusing on the effects of valence and other track-level audio features (i.e., features computed over entire musical pieces) on MECs.

Specifically, we compared features between tracks known to elicit MECs and a control set of tracks matched by artist, duration, and popularity. This experiment allowed us to collect large amounts of data. As a consequence, to promote transparency, we decided to clearly distinguish between confirmatory and exploratory analyses, as recommended by Dushoff et al. (2019). First, we conducted confirmatory analyses regarding the effect of the valence feature on MECs, hypothesising a difference in expressed valence between the two sets of tracks. Then, we conducted two exploratory analyses, to investigate the influence of other features on the occurrence of MECs, and to assess whether these features influenced the direction and magnitude of the difference in valence between both sets of tracks, with the aim to reveal whether or not differences in musical characteristics (i.e., different types of music) could be the reason behind conflicting effects of valence on MECs. We discuss these results and the advantages and limitations of our approach, and provide a new interpretation with reference to a theory of the personality correlates of musical preference (Rentfrow et al., 2011) introduced in Chapter 3, identifying relationships between

MECs and several dimensions capturing musical preference for *sophisticated* music and *intense* music.

4.2 Methods

4.2.1 Stimuli and features

Dataset

For this study, we used *ChiM*,[2] a dataset prepared by compiling every mention of a piece of music reported to elicit MECs in the literature reviewed in Chapter 2, following the methods described below.

In ChiM, pieces of music were only included if they were confirmed to elicit MECs in at least one listener. As a result, some pieces of music mentioned in the literature were intentionally omitted, if occurrences of MECs were not precisely assigned to individual pieces of music (e.g., Mas-Herrero et al., 2014), if no occurrences of MECs were recorded for specific pieces of music (e.g., Grewe et al., 2007), or if MECs occurred in response to stimuli combining music and other modalities (e.g., Strick et al., 2015). In the literature reviewed in Chapter 2, most mentions of pieces of music were included directly in the text or in the associated supplementary materials. For Bannister and Eerola (2018), data from the associated dataset (Bannister & Eerola, 2017) was also included.

The pieces of music were categorised according to whether each specific mention represented an anecdote by the authors, a participant report of MECs, an empirical verification of MECs, or a discussion of prior results. Artist names and song titles were harmonised across publications, most often following a basic search on Google or Wikipedia. Each mention of a piece of music was assigned a unique ID, and individual pieces of music were also assigned a unique ID in order to easily identify the ones which were mentioned several times in the literature, as indicated by an additional variable. Mentions of several movements from the same piece of music and otherwise duplicated mentions of pieces of music in the same publication were consolidated into a row with a single mention ID. Different performances or covers of a piece of music were assigned to the original composer, with a separate variable containing notes about further details on the performance. These pieces of music were assigned the same song ID with a distinct suffix. Finally, whenever a reasonable guess was possible, missing artist information was added. In other cases, missing information was indicated as such.

[2] ChiM version 1.0.0, available on Zenodo as a permanent archive, at https://doi.org/10.5281/zenodo.3950516

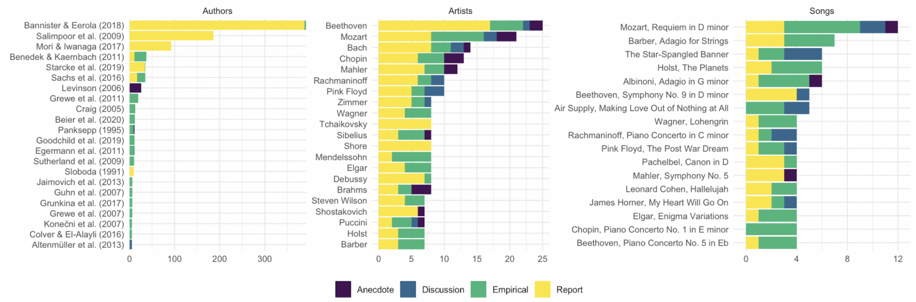

Figure 4.1: Count of pieces of music in version 1.0.0 of ChiM. The first plot shows the source of most items in ChiM, with more than 500 pieces of music originating from two articles only. The two other plots show the most frequent composers and pieces of music in ChiM. The colours in each bar represent whether a piece of music was an anecdote by the authors, a participant report of MECs, an empirical verification of MECs, or a discussion of prior results.

Version 1.0.0 of ChiM includes 988 mentions of music confirmed to induce MECs in at least one listener, which represents a much larger sample of music than the one explored in Chapter 3, and was therefore well-suited for computational analysis. As seen in Figure 4.1, the vast majority of ChiM consists of contributions from a few individual articles, and we can observe a slight over-representation of specific composers and pieces of music.

Features

Track-level audio features were collected using *spotifyr* (Thompson et al., 2020), an R package which enables pulling track information from the Spotify Web API. This allowed us to obtain, for most tracks, a range of features of interest, including *duration* (in milliseconds) and *popularity* (based on number and recency of plays), as well as nine track-level audio features: *acousticness* (confidence that the track is acoustic), *danceability* (based on tempo, rhythm stability, beat strength, and overall regularity), *energy* (based on dynamic range, perceived loudness, timbre, onset rate, and general entropy), *instrumentalness* (confidence that the track contains no vocal content), *liveness* (likelihood that the track was performed live), *loudness* (overall loudness in decibels), *speechiness* (presence of spoken words), *valence* (conveyed musical positiveness), and *tempo* (estimated in beats per minute). While Spotify does not share details about how these audio features are computed, they have been used successfully in previous research (e.g., Mas-Herrero et al., 2018; Melchiorre & Schedl, 2020).

4.2.2 Matching procedure

Chills source

We removed 136 duplicated tracks from ChiM, before looking up track information by sending API queries for strings containing the artist (or arranger/interpreter, as indicated in ChiM) and title of each track. The top result for each query was retained, and used to pull the features described above. Throughout this process, an additional 103 tracks were removed due to unavailability on Spotify or missing audio features, resulting in a dataset of 749 tracks which can cause MECs.

Matched source

Our analysis aimed to identify if specific track-level features were related to the occurrence of MECs in ChiM. Therefore, a control set of tracks which do not cause MECs was needed. Since it is impossible to assert that a specific track never causes MECs for anyone, we approximated the construction of this control set. More specifically, we compared features across tracks from the chills source with features in another set of tracks, matched as fairly as possible with the chills source by artist, duration, and popularity. This strict matching procedure ensured there were as few differences as possible between both sets of tracks other than their potential to elicit MECs. While it is possible that some tracks from the matched source could elicit MECs as well (see Chapter 3 and the discussion in this chapter for some limitations in our approach), it is unlikely that all of them would, meaning that any difference detected between the two sets of tracks should shed light on which factors affect the occurrence of MECs.

We decided to improve the matching approach used in Chapter 3 by implementing an algorithmic matching method. First, we gathered potential matches by getting the first 50 tracks for each of the first 50 albums returned by an API query for each artist represented in the tracks from the chills source. Then, we removed from these potential matches any track which was already present in the chills source, by comparing Spotify track IDs across the two sets. However, many duplicates remained across the two sets, since a piece of music on Spotify can have several distinct track IDs or slightly different titles. In order to mitigate this possibility, for each artist, we also removed from the potential matches any track with a title that had any number or any word of four letters or more (except the words *major* and *minor*) in common with tracks from the chills source. This process resulted in a pool of 205,717 potential matches.

Finally, we standardised track duration and popularity across tracks from the chills source and the potential matches. For each artist, the potential matches with the shortest Euclidean distance from each track from the chills source for

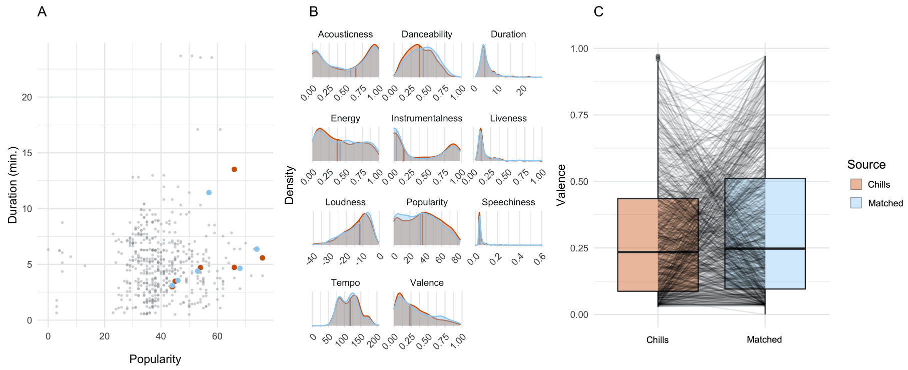

Figure 4.2: **A.** Example of the matching procedure, using Pink Floyd tracks. Tracks from ChiM are shown in orange, and potential matches gathered with the Spotify Web API are shown in grey. Potential matches with the shortest Euclidean distance from each track from the chills source, in terms of duration and popularity, were selected as matches, shown in blue. **B.** Densities and median values of audio features and metadata for the 722 resulting pairs of tracks from the chills and matched sources. **C.** Boxplots showing valence for the 722 pairs of tracks, with lines linking valence scores for each individual pair.

these two standardised features were retained as the best matches. Audio features were pulled for these 749 matches, but were missing for 10 of them. A further 17 matches were considered as outliers and removed for having Euclidean distances larger than three standard deviations from the mean, resulting in 722 pairs of tracks from the chills and matched sources. The matching procedure is illustrated in Figure 4.2, and the full list of tracks from the chills and matched sources is included in the supplementary materials of the published article corresponding to this chapter (de Fleurian & Pearce, 2021).[3]

4.3 Analysis

The confirmatory analysis consisted of assessing whether there was a difference in valence between tracks from the chills and matched sources. We ran a logistic regression for the effect of the valence feature on track source (chills vs. matched). The presence of influential data points was checked with leave-one-out diagnostics, with the plan to run another logistic regression excluding data points that would, if left out, affect the slope by at least half of its original absolute value.

The exploratory analyses were twofold. First, we assessed whether there was a difference between tracks from the chills and matched sources in the

[3]This list of tracks is not included as an appendix to the present thesis due to its large size, better suited for a CSV-formatted file.

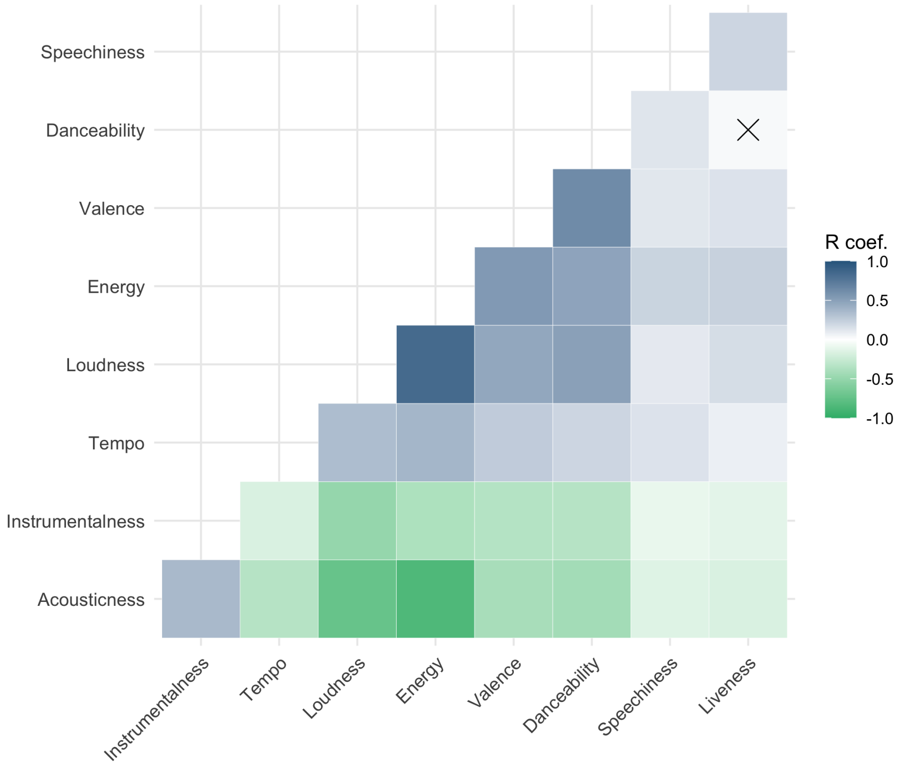

Figure 4.3: Correlation matrix, showing the high degree of collinearity between each feature. Positive correlations are shown in blue and negative correlations in green. Colour saturation corresponds to the magnitude of each correlation, and the only non-significant correlation is indicated by a cross in the corresponding cell.

nine audio features described earlier. Due to high collinearity between these features (see Figure 4.3), we first ran a principal component analysis (PCA), a method which projects each data point into a new dimensional space, the axes of which are called principal components, and allows for dimensionality reduction when retaining the first few components, which account for as much variance in the data as possible. PCA was run with centring and scaling, resulting in two principal components with eigenvalues above one—a common threshold to decide which components to retain based on how much variance they explain in the original data. We ran a logistic regression for the effects of these two principal components on track source (chills vs. matched), checking influential data points with leave-one-out diagnostics as described above.

Second, we assessed whether the audio features of the tracks from the chills source had an effect on the *difference* in valence between tracks from the chills and matched sources, which could possibly mean that different types of MECs arise in response to different auditory characteristics. As described above, we first ran a PCA due to collinearity, before running a linear regression for the effects of the two resulting principal components on the difference in valence between the tracks from the chills and matched sources, using leave-one-out diagnostics, and checking homoscedasticity and normality of residuals with residuals plots.

Finally, to check the robustness of the matching procedure, we conducted Wilcoxon signed-rank tests to compare duration and popularity between tracks from the chills and matched sources, expecting no significant differences in duration and popularity. Nonparametric tests were chosen due to the fact that duration and popularity did not follow a normal distribution (see Figure 4.2). Since there were some differences between both sets of tracks (see Section 4.4), we conducted mediation analyses, using the nonparametric bootstrap with 5000 Monte Carlo draws, as implemented in the *mediation* R package (Tingley et al., 2014), to check if potential effects of the valence feature on track source were mediated by track duration and popularity. Mediation analysis enables the separation of the total effect into average causal mediation effect (ACME, the indirect effect of the independent variable on the dependent variable that goes through the mediator) and average direct effect (ADE, the direct effect of the independent variable on the dependent variable). In addition, to mitigate this weakness of the matching procedure, all the analyses described above were replicated a total of 10 times, using a different set of matched sources, each comprising one of the 10 tracks with the shortest Euclidean distance from each track from the chills source (i.e., shortest Euclidean distance for iteration #1, second shortest for iteration #2, etc.)

4.4 Results

Effect of valence on track source

A logistic regression model yielded a significant fit ($\chi^2(1) = 6.33$, $p = .012$, Nagelkerke $R^2 = .006$), revealing a significant effect of the valence feature on track source ($b = 0.54$, $Z = 2.51$, $p = .012$), with the valence of tracks from the chills source being lower than that of tracks from the matched source by 0.033 on a 0–1 scale (see Figure 4.2). This effect remained significant across all 10 iterations of the analysis (mean valence difference = 0.042, $SD = 0.009$). Results for the 10 iterations are shown in Table 4.1.

Table 4.1: Effect of valence on track source

	Model fit			Valence		
Iteration	χ^2	p	Nagelkerke R^2	b	Z	p
1	6.33	.012	.006	0.54	2.51	.012
2	12.04	$< .001$	.011	0.75	3.45	$< .001$
3	8.99	.003	.008	0.65	2.99	.003
4	5.13	.023	.005	0.51	2.26	.024
5	13.13	$< .001$	.012	0.80	3.60	$< .001$
6	6.42	.011	.006	0.56	2.53	.012
7	11.72	$< .001$	.011	0.75	3.40	$< .001$
8	15.46	$< .001$	.014	0.83	3.91	$< .001$
9	15.88	$< .001$	.015	0.86	3.96	$< .001$
10	14.98	$< .001$	.014	0.84	3.84	$< .001$

Mediating effects of duration and popularity

A Wilcoxon signed-rank test revealed no significant difference in duration ($V = 137199$, $p = .232$) and a significant difference in popularity ($V = 134593$, $p = .003$) between tracks from the chills and matched sources (higher for the chills source by 2.90 on a 1–100 scale), suggesting that the matching procedure did not result in an optimal set of matched tracks. The difference in popularity remained significant in all 10 iterations of the analysis, while the difference in duration became significant in the fourth as well as the last five iterations of the analysis, presumably due to the increasing Euclidean distance between tracks from the chills source and each successive set of matched sources (see Table 4.2).

Table 4.2: Difference in duration and popularity between track sources

	Duration		Popularity	
Iteration	V	p	V	p
1	137199	.232	134593	.003
2	142583	.051	145166	$< .001$
3	142185	.060	150597	$< .001$
4	146264	.005	152310	$< .001$
5	140323	.059	154629	$< .001$
6	141824	.026	159205	$< .001$
7	145974	.005	159076	$< .001$
8	144000	.013	162740	$< .001$
9	143062	.010	163894	$< .001$
10	144711	.004	165209	$< .001$

To assess whether duration and popularity mediated the effect of the valence feature on track source as reported above, we conducted two separate causal mediation analyses. For duration, the average causal mediation effect was not significant ($ACME = -.006$, $p = .716$) and the average direct effect was significant ($ADE = -.128$, $p = .021$), suggesting that duration did not mediate the effect of valence on track source. For popularity, both average effects were significant ($ACME = .037$, $p = .002$; $ADE = -.171$, $p < .001$), suggesting that popularity partially, but not fully, mediated the effect of valence on track source. These results remained stable across the 10 iterations of the analysis, except for duration, which partially mediated the effect of valence on track source in the last iteration (see Table 4.3).

Table 4.3: Mediation analyses for the effect of valence on track source

	Duration				**Popularity**			
Iteration	$ACME$	p	ADE	p	$ACME$	p	ADE	p
1	-.006	.716	-.128	.021	.037	.002	-.171	$< .001$
2	-.010	.503	-.173	.002	.049	$< .001$	-.232	$< .001$
3	-.012	.434	-.147	.004	.063	$< .001$	-.223	.001
4	-.028	.103	-.100	.095	.067	$< .001$	-.196	$< .001$
5	-.012	.419	-.184	$< .001$	.073	$< .001$	-.270	$< .001$
6	-.020	.232	-.118	.034	.072	$< .001$	-.209	$< .001$
7	-.019	.209	-.164	.002	.073	$< .001$	-.256	$< .001$
8	-.016	.267	-.189	$< .001$	.075	$< .001$	-.281	$< .001$
9	-.022	.145	-.188	.001	.073	$< .001$	-.285	$< .001$
10	-.032	.046	-.174	.002	.081	$< .001$	-.288	$< .001$

In this case, the mediation analyses each involved a linear regression (for the effect of valence on duration/popularity) and a logistic regression (for the effects of valence and duration/popularity on track source). It is worth noting that for the linear models, some assumptions (homoscedasticity and normality of residuals) were violated, as shown in Figure 4.4. This was most likely due to the distribution of valence, duration, and popularity in our data (see Figure 4.2). To confirm the results of the mediation analyses, we ran them again on a reduced (keeping only tracks with non-zero popularity due to this feature being zero-inflated) and transformed dataset (square root for popularity and log for valence and duration). These re-analyses did not fully eliminate the violations of assumptions for linear regression, but did replicate the findings presented above (see Table 4.4).

Effects of audio features on track source

We ran a PCA to reduce collinearity in the nine audio features. We retained two principal components with eigenvalues higher than one, accounting for 56.4% of cumulative proportion of variance explained (see supplementary materials discussed earlier in de Fleurian & Pearce, 2021, for the values for each set of tracks). The first component featured high positive loadings (greater than .2) for energy, loudness, valence, danceability, and tempo, and high negative loadings

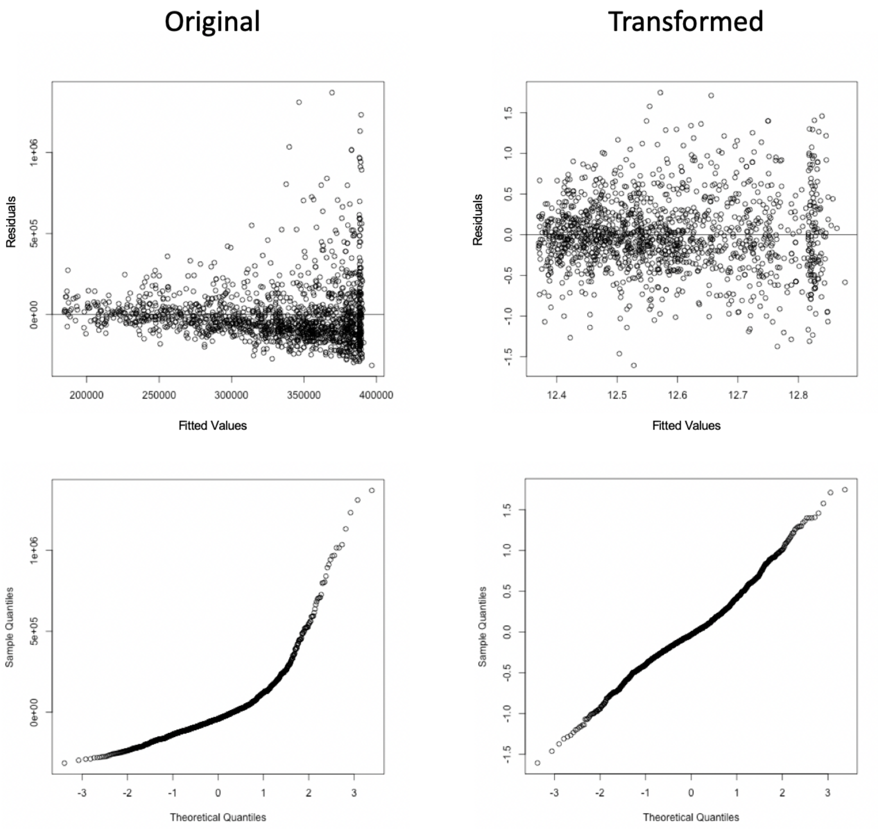

Figure 4.4: Diagnostic plots for a linear regression of the effect of valence on track duration. Plots on the left show a violation of homoscedasticity in the top residuals plot, and a violation of normality in the bottom Normal Q-Q plot when using untransformed variables. Plots on the right show that, when using log-transformed variables for valence and track duration instead, the assumptions for linear regression were much better fulfilled. Note that these improvements were not as noticeable for all models.

(lower than -.2) for acousticness and instrumentalness. The second component featured high positive loadings for liveness and speechiness, and a high negative loading for danceability. The number of retained principal components and their associated loadings were consistent across all 10 iterations of the analysis (besides occasional but systematic sign differences, which are expected when conducting several PCAs—see Table 4.5).

A logistic regression model yielded a significant fit ($\chi^2(2) = 6.47$, $p = .039$, Nagelkerke $R^2 = .006$), revealing a significant effect of the first component on track source ($b = 0.06$, $Z = 2.34$, $p = .019$) and no significant effect of the second component ($b = 0.05$, $Z = 0.98$, $p = .328$), showing that tracks from the chills source had lower scores than tracks from the matched source

Table 4.4: Mediation re-analyses for the effect of valence on track source

	Duration				Popularity			
Iteration	*ACME*	*p*	*ADE*	*p*	*ACME*	*p*	*ADE*	*p*
1	-.002	.656	-.023	.109	.010	.001	-.036	.010
2	-.005	.304	-.034	.024	.013	$<$ .001	-.052	$<$ .001
3	-.006	.136	-.019	.174	.017	$<$ .001	-.043	.002
4	-.009	.086	-.024	.102	.020	$<$ .001	-.052	$<$ .001
5	-.006	.217	-.030	.032	.020	$<$ .001	-.056	$<$ .001
6	-.007	.094	-.017	.242	.021	$<$ .001	-.045	$<$ .001
7	-.008	.051	-.028	.058	.023	$<$ .001	-.059	$<$ .001
8	-.005	.198	-.028	.049	.023	$<$ .001	-.056	$<$ .001
9	-.008	.070	-.036	.014	.021	$<$ .001	-.065	$<$ .001
10	-.010	.026	-.036	.016	.026	$<$ .001	-.072	$<$ .001

on the first component (i.e., tracks from the chills source featured lower energy, loudness, valence, danceability, and tempo, as well as higher acousticness and instrumentalness—see Figure 4.5). The model fit remained significant in all but one iteration of the analysis ($\chi^2(2) = 5.33$, $p = .070$, Nagelkerke $R^2 = .005$), the effect of the first component remained significant in all iterations, and the effect of the second component became significant in four iterations, highlighting that in some cases, tracks from the chills source had lower scores than tracks from the matched source on the second component (i.e., low liveness and speechiness, as well as high danceability). It is worth noting that in the seventh iteration of the analysis, there were two influential data points (as described in Section 4.3). For this iteration, we ran the model both with and without the influential data points, leading to similar results in both cases (see Table 4.6).

Effects of audio features on difference between sources

We ran a PCA on the nine audio features of the tracks from the chills source only (as opposed to both track sources in Section 4.4), to assess if properties of the tracks from the chills source could predict the direction and magnitude of the difference in the valence feature between both track sources. We retained two principal components with eigenvalues higher than one, accounting for 55.6% of cumulative proportion of variance explained. Both components featured similar loadings as in the previous section. The number of retained principal components and their associated loadings were consistent across all 10 iterations of the analysis (see Table 4.7).

A multiple linear regression model yielded a significant fit ($F(2, 719) = 63.9$, $p < .001$, adjusted $R^2 = .149$), revealing a significant effect for both the first component ($\beta = .062$, $p < .001$) and the second component ($\beta = .039$, $p < .001$), suggesting that tracks from the chills source with higher scores on

Table 4.5: PCA on audio features for all tracks

		Audio feature loadings								
PC	I.	Temp.	Loud.	Val.	Danc.	Ener.	Acou.	Inst.	Spee.	Live.
1	1	.242	.439	.353	.346	.457	-.419	-.291	.132	.134
	2	.222	.444	.356	.346	.459	-.422	-.298	.128	.118
	3	.233	.438	.350	.343	.458	-.421	-.289	.151	.142
	4	.220	.447	.351	.341	.461	-.428	-.300	.112	.117
	5	.221	.442	.356	.343	.461	-.422	-.297	.139	.117
	6	.243	.439	.354	.338	.456	-.420	-.292	.153	.124
	7	.231	.439	.346	.341	.459	-.425	-.301	.151	.115
	8	.241	.437	.352	.348	.453	-.419	-.302	.135	.125
	9	.220	.441	.356	.348	.456	-.421	-.293	.151	.124
	10	.241	.438	.351	.351	.456	-.422	-.292	.130	.122
2	1	.119	-.083	-.178	-.275	.048	-.001	.105	.645	.665
	2	.158	.078	.080	.148	.003	-.071	.046	-.660	-.706
	3	.138	-.029	-.293	-.365	.081	-.051	.098	.548	.665
	4	.075	-.052	-.240	-.300	.062	-.055	.020	.581	.706
	5	.050	.019	.219	.324	-.063	.029	.013	-.567	-.719
	6	.007	-.020	-.261	-.359	.076	-.048	.007	.498	.739
	7	.043	-.010	-.276	-.363	.048	-.038	-.045	.543	.699
	8	.037	.078	.150	.202	.017	-.046	.056	-.644	-.714
	9	.052	-.075	-.173	-.267	.020	.012	.035	.612	.717
	10	.088	.100	.101	.162	.007	-.051	.033	-.654	-.717

Note. PC = Principal component, I. = Iteration, Temp. = Tempo, Loud. = Loudness, Val. = Valence, Danc. = Danceability, Ener. = Energy, Acou. = Acousticness, Inst. = Instrumentalness, Spee. = Speechiness, Live. = Liveness.

these components were more likely to be happier than their associated tracks from the matched source, and vice versa. These effects held for all 10 iterations of the analysis (see Table 4.8).

4.5 Discussion

Results

In this experiment, we compared track-level audio features between tracks taken from ChiM, a dataset of pieces of music known to elicit MECs, and several sets of tracks algorithmically matched by artist, duration, and popularity.

We compared the valence feature between tracks from the chills and matched sources, and found that valence was, on average, slightly lower in the chills source. This echoes previous findings that MECs are more frequently associated with perceived sadness (Panksepp, 1995), as opposed to perceived happiness (Grewe et al., 2011). The matching procedure resulted in a small difference in valence between track sources, but it is worth noting that overall, the distribution of valence in ChiM is highly positively skewed, whereas it is relatively uniform across all tracks on Spotify, as seen in the documentation for the valence feature in the Web API.[4] In other words, an effect of valence was identified despite the application of a strict matching procedure, which most likely resulted in high

[4]https://developer.spotify.com/documentation/web-api

Table 4.6: Effects of first two principal components on track source

	Model fit			**Component 1**			**Component 2**		
Iteration	χ^2	p	Nagelkerke R^2	b	Z	p	b	Z	p
1	6.47	.039	.006	0.06	2.34	.019	0.05	0.98	.328
2	10.39	.006	.010	0.07	2.42	.016	-0.11	-2.10	.036
3	6.57	.038	.006	0.07	2.51	.012	0.03	0.51	.612
4	13.04	.001	.012	0.07	2.63	.008	0.12	2.44	.015
5	9.00	.011	.008	0.07	2.76	.006	-0.06	-1.17	.243
6	5.33	.070	.005	0.06	2.15	.032	0.04	0.83	.404
7a	6.71	.035	.006	0.07	2.56	.010	0.02	0.37	.714
7b	6.96	.031	.006	0.07	2.61	.009	0.02	0.40	.687
8	13.28	.001	.012	0.07	2.74	.006	-0.12	-2.37	.018
9	7.20	.027	.008	0.07	2.61	.009	0.03	0.60	.551
10	12.53	.002	.012	0.07	2.69	.007	-0.12	-2.26	.024

Note. The analysis for iteration 7 was conducted with (7a) and without (7b) influential data points.

similarity between tracks from the chills and matched sources. If control tracks had been selected randomly instead, most tracks from the chills source would have had a much lower valence by comparison.

When taking all audio features into consideration, we found that tracks from the chills source were characterised by smaller values on a component linked with high energy, loudness, valence, danceability, and tempo, as well as low acousticness and instrumentalness, meaning that overall, tracks from the chills source were sadder, slower, less intense, and more instrumental than tracks from the matched source. In a few occasions, the chills source was also characterised by smaller values on a second component linked with high liveness and speechiness, as well as low danceability, therefore suggesting that music that causes MECs may be less likely to include spoken words and to feature a live audience, although these results were less robust than those for the first component. These findings can be interpreted with reference to an influential theory of the personality correlates of musical preference (Rentfrow et al., 2011), which we briefly mentioned in Chapter 3. While music preference tests are not necessarily best suited to studying the impact of stylistic preference on the occurrence of MECs, as discussed in said chapter, it is worth noting that the musical characteristics we identified in the present study strongly matched *sophisticated* music, which tends to be relaxing, quiet, non-danceable, slow, non-electric, and instrumental (Rentfrow et al., 2012), suggesting that tracks from the chills source were more sophisticated than tracks from the matched source in our analysis. Interestingly, preference for sophisticated music is associated with openness to experience (Schäfer & Mehlhorn, 2017), a personality characteristic strongly linked to the experience of MECs (see Chapter 2).

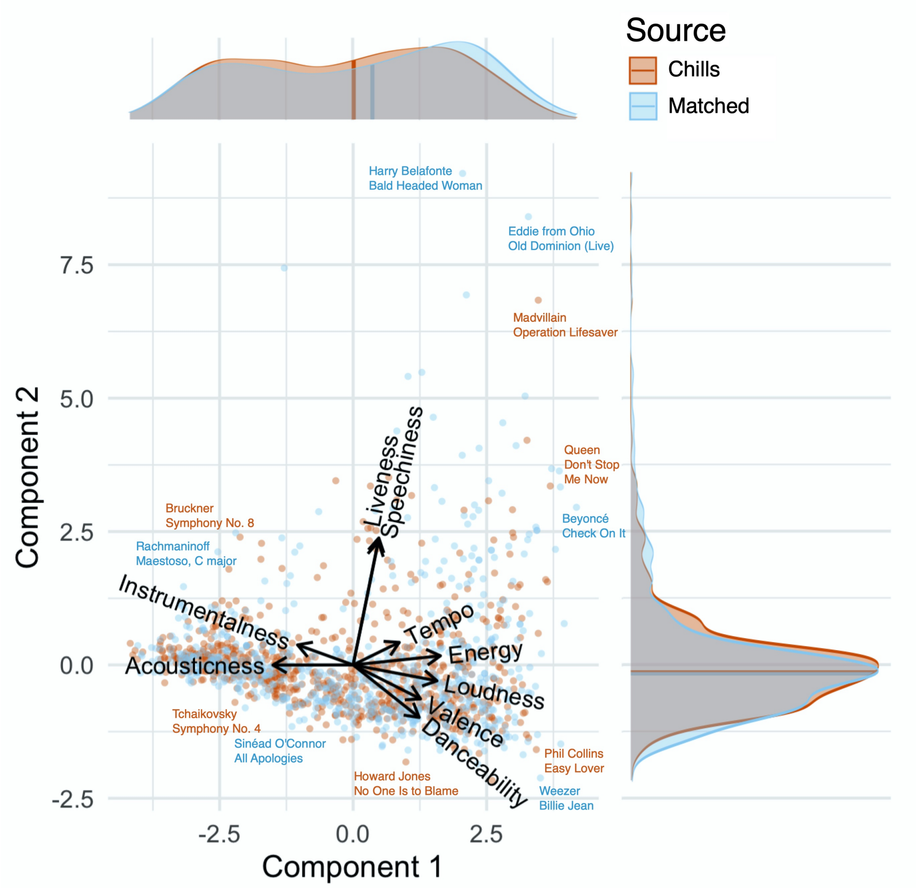

Figure 4.5: Biplot of tracks from the chills and matched sources for the first iteration of the analysis. Tracks are mapped onto the first two components obtained with PCA. Some example tracks are shown for various combinations of component values. Densities and median values for tracks from the chills and matched sources are shown in marginal plots, revealing a difference on Component 1. Audio feature loadings are shown as vectors, illustrating the high degree of collinearity between some features.

We also examined whether the audio features of tracks from the chills source related to the difference in valence between the chills and matched sources. We found that tracks with higher energy, loudness, valence, liveness, and speechiness, as well as lower acousticness and instrumentalness, were more likely to feature higher expressed happiness than their associated matched tracks, and vice versa. While these results are partly explained by valence loading on the first component obtained with PCA, the involvement of other audio features suggests a potential interpretation. Using the same classification as above (Rentfrow et al., 2012), it becomes apparent that, on average, *sophisticated* tracks from the chills source were sadder than their matched tracks, and *intense* tracks from the chills source

Table 4.7: PCA on audio features for tracks from the chills source only

		Audio feature loadings								
PC	I.	Temp.	Loud.	Val.	Danc.	Ener.	Acou.	Inst.	Spee.	Live.
1	1	.251	.440	.345	.339	.463	-.419	-.283	.155	.126
	2	.249	.440	.346	.336	.463	-.420	-.283	.158	.127
	3	.249	.440	.347	.335	.463	-.421	-.283	.157	.125
	4	.247	.441	.346	.334	.464	-.421	-.282	.153	.129
	5	.249	.440	.346	.337	.463	-.420	-.281	.156	.129
	6	.253	.439	.344	.337	.463	-.420	-.282	.157	.125
	7	.250	.440	.346	.335	.463	-.421	-.282	.158	.126
	8	.250	.440	.346	.335	.463	-.421	-.282	.158	.126
	9	.249	.439	.346	.334	.463	-.421	-.285	.159	.126
	10	.250	.438	.345	.338	.462	-.420	-.286	.155	.127
2	1	.014	.088	.149	.265	-.020	-.007	-.037	-.650	-.690
	2	.008	.083	.164	.283	-.030	.004	-.049	-.646	-.682
	3	.009	.081	.164	.281	-.029	.004	-.047	-.646	-.684
	4	.005	.084	.162	.284	-.030	.005	-.049	-.654	-.674
	5	.025	.086	.148	.273	-.024	-.005	-.040	-.648	-.688
	6	.016	.086	.157	.269	-.024	-.004	-.040	-.647	-.689
	7	.015	.088	.152	.274	-.025	-.001	-.047	-.646	-.688
	8	.015	.088	.152	.274	-.025	-.001	-.047	-.646	-.688
	9	.009	.084	.162	.277	-.028	.000	-.052	-.646	-.685
	10	.011	.082	.160	.277	-.031	.000	-.048	-.650	-.682

Note. PC = Principal component, I. = Iteration, Temp. = Tempo, Loud. = Loudness, Val. = Valence, Danc. = Danceability, Ener. = Energy, Acou. = Acousticness, Inst. = Instrumentalness, Spee. = Speechiness, Live. = Liveness.

(i.e., non-relaxing, loud, electric, and featuring raspy or yelling voice) were happier than their matched tracks. In other words, the directionality of the difference in valence between both sets of tracks was determined by the degree to which tracks from the chills source were relaxing, quiet, and instrumental. As is the case with sophisticated music, intense music is linked with openness to experience (Schäfer & Mehlhorn, 2017), a known personality correlate of MECs. Interestingly, these results provide some support for the possibility that different types of MECs are elicited by different types of feelings and affective states expressed or evoked by music (Bannister, 2019; Maruskin et al., 2012), and for the possible presence of several pathways for the experience of MECs (see Chapter 2).

Limitations

We discussed the rationale and some limitations of the matching procedure in Section 4.2.2. Notably, despite our best efforts, tracks from the matched source were slightly less popular than their counterparts. This could be due to MECs being more likely in popular tracks, either because the repeated listening associated with track popularity contributes to the elicitation of MECs, or because the ability to elicit MECs contributes to tracks becoming popular. In our opinion, however, this difference is most likely due to a bias towards popular tracks when reporting music which causes MECs. It should be possible to reduce

Table 4.8: Effects of components on difference in valence between sources

	Model fit			Component 1		Component 2	
Iteration	F	p	Adjusted R^2	β	p	β	p
1	63.9	$< .001$	.149	0.062	$< .001$	0.039	$< .001$
2	62.5	$< .001$	.145	0.059	$< .001$	0.040	$< .001$
3	65.2	$< .001$	.151	0.063	$< .001$	0.033	.002
4	64.8	$< .001$	.150	0.059	$< .001$	0.031	.002
5	55.7	$< .001$	.132	0.057	$< .001$	0.028	.007
6	47.5	$< .001$	.115	0.052	$< .001$	0.026	.012
7	51.8	$< .001$	.124	0.055	$< .001$	0.033	.002
8	51.4	$< .001$	.123	0.057	$< .001$	0.035	.002
9	51.7	$< .001$	.124	0.058	$< .001$	0.033	.003
10	55.2	$< .001$	.131	0.058	$< .001$	0.035	.001

this difference by picking matched tracks from a larger set of potential matches, but our current methods did not allow this due to limits on the rate of Spotify API requests and the number of records per request. As a result, we found that popularity partially, but not fully, mediated the difference in valence between both sets of tracks. This could be interpreted as sad songs being more popular (see Gulmatico et al., 2022), and in turn, popular songs being more likely to cause MECs. However, the mediation analyses were not part of the planned primary analysis, but rather only intended as a procedural check, and we suspect the identified mediating effects were largely due to the idiosyncrasies of our data. Moreover, there was still a residual direct effect of valence in our analysis, which should not be overlooked. Nonetheless, we attempted to address any limitations in the matching procedure by conducting 10 iterations of the full analysis with different sets of matched sources, which led to consistent results across iterations.

There were other limitations to our approach. ChiM does not report the exact version of the pieces of music that elicit MECs, which could have had some impact on the audio characteristics of the tracks from the chills source. Then, some tracks from the chills source might have been present in the matched source, despite our efforts to limit this possibility (see Section 4.2.2). In general, apart from a few sanity checks, we considered the lack of manual verification of our data as an acceptable trade-off for the large size of the dataset, which made a robust computational analysis possible. Another issue is the lack of transparency about how Spotify computes audio features. Again, we accepted this trade-off which allowed us to collect large amounts of audio data and metadata through API queries.

More importantly, and as initially suspected following the results presented in Chapter 3, effect sizes were small for most of our results. One possible explanation is that we tried to be as fair as possible with the matching procedure, which might have drastically reduced the differences in audio features between tracks from the chills and matched sources—effect sizes would probably have been more pronounced if we randomly selected matches, but this process would have introduced noise due to confounding differences between both track sources. Also, as highlighted in Chapter 3 and Section 4.2.2, it is possible that some tracks from the matched source also had the ability to elicit MECs. Finally, MECs are a localised phenomenon, and it is fully expected that track-level features would not capture local changes in acoustic and structural features, therefore limiting the explanatory power of our approach. However, we believe that the consistency of the results across several iterations of the analysis on different sets of matched sources yielded robust and interpretable findings, despite their small effect size.

Conclusion

We conducted a corpus analysis of audio characteristics of music known to elicit MECs and identified that such music was sadder than music matched by artist, duration, and popularity. Exploratory analyses revealed that tracks from the chills source were also slower, less intense, and more instrumental than tracks from the matched source on average, and tended to possess the characteristics of sophisticated music. Moreover, within the chills source, we identified a possible relationship between valence and type of music, with sophisticated music tending to be sadder than tracks from the matched source, and intense music tending to be happier. Taken together, these findings provide further support for an effect of perceived valence on the occurrence of MECs, and for a possible pathway for the experience of MECs involving the psychological mechanism of emotional contagion and the process of being moved. However, they do not exclude the possibility of separate pathways (see Chapter 2).

Overall, these results showed that, for research on MECs, computational methods have a great and largely untapped potential to complement behavioural studies. More specifically, such methods were able to identify that, when looking at track-level characteristics, music that causes MECs differed in musical content from other music, which motivates further computational investigation into the local acoustic and musical elicitors of MECs, as conducted in Chapter 5. Regarding future directions in research on MECs and valence, some questions remain. Notably, it would be worth investigating the effect of local changes in valence on MECs as opposed to track-level aggregates, which would provide further insight into the interaction between global context (e.g., music that is happy or sad, sophisticated or intense) and local events (e.g., specific happy or sad passages).

Chapter 5

Musical expectation

As opposed to the work presented in Chapter 3, the study discussed in Chapter 4 led to the identification of differences between tracks previously reported as being able to cause MECs and tracks matched to those by artist, duration, and popularity, lending support to the intuition that such differences are more easily detected using computational methods on a large corpus of music. The work presented so far has considered each piece of music as a single data point, i.e., each piece of music was characterised by a one-dimensional vector of extracted features and behavioural responses. But, as discussed in Chapter 2, MECs correspond to transient events occurring dynamically at particular points within a piece of music.

This chapter presents a computational analysis aimed at modelling the onset of MECs based on acoustic and musical characteristics (see Chapter 2). We used the dataset of pieces of music initiated in Chapter 3, labelled with onsets of MECs, and extracted features corresponding to previously identified acoustic and musical elicitors of MECs, as well as features capturing widely hypothesised elicitors of MECs, such as musical expectation. In the first part of the present study, we ran a series of permutation tests for each feature around the onsets of MECs, confirming in a systematic way, and at a much larger scale, the correlational effects that have been identified in previous research. In the second part, we compared the performance of two classification approaches, by training two different types of models on excerpts centred around the onsets of MECs and randomly selected excerpts from the same pieces of music, and testing these models in an automatic MEC onset detection task, resulting in the findings that the onsets of MECs could be predicted better than chance, and that musical expectation was the most effective predictor of MECs.

5.1 Introduction

A major motivation for the present study was to compare the predictive performance of the previously identified elicitors of MECs. In order to do so, it is worth expanding on Chapter 2 by briefly reviewing these elicitors with a view to selecting a list of appropriate features for the analyses presented in this chapter.

Acoustic elicitors

Acoustic elicitors of MECs refer to low-level properties of the auditory signal. In early research on MECs, Sloboda (1991) identified a relationship with sudden dynamic changes, by analysing music scores for passages reported to elicit MECs by survey participants. This relationship was confirmed empirically in subsequent research (Auricchio, 2017; Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Guhn et al., 2007; Honda et al., 2020; Nagel et al., 2008; Polo, 2017), in which loudness was extracted (or manually inspected from music scores in the case of Guhn et al., 2007) around the onset of MECs experienced by participants when listening to music in a lab environment, which they reported either retrospectively, or continuously by pressing a button or moving a slider.

Loudness is by far the most documented acoustic correlate of MECs, but other relationships were also identified using similar methods, suggesting that occurrences of MECs tend to correlate with increases in roughness, dissonance, or fluctuation strength (Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Nagel et al., 2008; Park et al., 2019), increases in sharpness or brightness (Bannister & Eerola, 2018; Beier et al., 2020; Grewe et al., 2007; Honda et al., 2020), high spectral centroid and spectral flux (Bannister & Eerola, 2018), high event density (Bannister & Eerola, 2018; Nagel et al., 2008; Polo, 2017), or expansion of the frequency range in a high or low register (Guhn et al., 2007; Polo, 2017).

In a recent study, Bannister (2020c) experimentally manipulated loudness and brightness in two musical passages that had elicited MECs in previous research (Bannister & Eerola, 2018). It was found that in one of the musical passages, MECs were experienced more frequently if loudness was increased, or if brightness was decreased (in contradiction with previous findings), therefore demonstrating a causal effect of loudness and brightness on MECs, as opposed to the correlational findings discussed above.

Musical elicitors

Musical elicitors of MECs refer to high-level properties of the musical structure. Sloboda (1991), in the same study discussed above, identified that musical passages causing MECs also included new or unprepared harmonies, sudden textural changes, melodic appoggiaturas, enharmonic changes, specific melodic or harmonic sequences, or prominent events arriving earlier than prepared for, in decreasing order of frequency.

These self-reported effects of melodic and harmonic properties on MECs were confirmed in subsequent survey-based and empirical research (Auricchio, 2017; Bannister, 2020b; Bannister & Eerola, 2018; Guhn et al., 2007; Mlejnek, 2013; Schurtz et al., 2012), notably through the identification of an effect of structural transitions and alterations such as changes in tonality (Bannister, 2020b). Rhythmic properties (Schurtz et al., 2012; Solberg & Dibben, 2019) and vocals (Bannister, 2020b; Schurtz et al., 2012) were also found to be involved, although there is a lack of specificity regarding which exact properties were associated with MECs.

Additional findings revealed effects of crescendi, build-ups, and climaxes (Auricchio, 2017; Bannister, 2020b; Bannister & Eerola, 2018; Panksepp, 1995; Polo, 2017; Solberg & Dibben, 2019), as well as textural changes (Auricchio, 2017; Polo, 2017; Sloboda, 1991; Solberg & Dibben, 2019), notably through the entrance of new instruments or the interplay between solo and background instruments (Auricchio, 2017; Bannister, 2020b; Bannister & Eerola, 2018; Goodchild et al., 2019; Guhn et al., 2007; Mlejnek, 2013).

Emotional elicitors

It is also worth mentioning emotional elicitors of MECs, which refer to subjectively perceived valence, emotionality, and meaning in music (see Chapter 2). These characteristics of musical stimuli are difficult to quantify precisely and objectively, especially as continuous features, since they rely on some degree of subjective interpretation. Acoustic and musical elicitors, however, refer to properties of the auditory signal and of the musical structure that do not exclusively rely on subjective judgements.

For this reason, emotional elicitors were not considered in the present study, but we acknowledge that they are considered potent elicitors of MECs, and deserve further attention in future research.

Expectation and chills

As discussed in Chapter 2, findings about musical elicitors have often been placed in the context of a hypothesised effect of musical expectation on MECs (L. Harrison & Loui, 2014; Huron, 2006; Huron & Margulis, 2010; Juslin, 2013; Juslin & Västfjäll, 2008; McDermott, 2012; Mencke et al., 2019; Pearce & Wiggins, 2012; Salimpoor et al., 2011; Sloboda, 1991), positing that most of the musical elicitors listed above could be related to violations of expectation, and in turn, to experiences of MECs.

An effect of musical expectation on emotional responses to music has long been hypothesised (Hanslick, 1854; Meyer, 1956), and has since been confirmed in empirical studies (Cheung et al., 2019; Egermann et al., 2013; Gold et al., 2019; Huron, 2006; Juslin, 2013; Sauvé et al., 2018; Steinbeis et al., 2006). The effect of expectation on MECs, however, remains untested.

Automatic detection of chills

While the findings presented above represent a wide range of elicitors which are associated with MECs, most of these findings are qualitative or correlational in nature. In addition, elicitors were often identified by subjective analysis of music scores, and based on a relatively small amount of data. Despite these limitations, the degree of consensus for their effects on MECs suggests that automatic detection of MECs based on these elicitors might be possible.

The predictive modelling of continuous responses such as MECs, in addition to having received little attention in prior research (Eerola, 2018), requires large amounts of data, given the complexity in how acoustic and musical characteristics vary over time and interact with each other. This is especially relevant when also trying to investigate the relative influence of each individual elicitor on the occurrence of MECs.

While the ChiM dataset used in Chapter 4 is suitably large for such a task, it lacks consistent information about the exact versions of the pieces of music which elicit MECs, as well as precise information about the timing of MEC onsets. Using the dataset introduced in Chapter 3 addresses both of these challenges.

Objectives

The present study aimed to extract and process features representing acoustic and musical elicitors of MECs from an empirical dataset of onsets of MECs, to conduct a robust analysis of the effects of these elicitors on the occurrence of MECs, expecting these effects to replicate previous findings, and to uncover new evidence about an effect of musical expectation on MECs.

In addition, this chapter reports the design of a computational system for the automatic detection of MECs, which we expected to perform better than chance when predicting onsets of MECs based on acoustic and musical features. This system allowed the investigation of feature importance when predicting MEC onsets, through which we expected to observe a large influence of musical expectation. The models were cross-validated as a part of the training process, therefore reducing uncertainty in the findings due to the exploratory nature of this part of the present study.

First, we describe the additional data obtained through the survey study described in Chapter 3. Then, we outline a set of features that broadly covered the range of elicitors discussed above. Finally, we detail the construction of analyses and models with the aim to provide a large-scale replication of current findings about the elicitors of MECs, to rank their importance in the elicitation of MECs, and to test the hypothesised effect of musical expectation on MECs.

5.2 Methods

5.2.1 Stimulus selection

Dataset

In order to collect a sufficient amount of training samples, considering the large number of features planned for model training, we left the survey study described in Chapter 3 running on Qualtrics (Qualtrics, Provo, UT) between February 2018 and April 2020. We collected a large number of self-reports of onsets of MECs from 2069 participants (including the 221 participants from Chapter 3) ranging in age from 18 to 77 years (M = 23.6 years, SD = 8.8 years), and originating from a wide range of geographical areas (62 % North America, 25% Europe, 8% Asia, 3% Oceania, 1% Africa, 1% South America).

The resulting data required extensive manual cleaning. First, we removed entries from participants who abandoned the survey study before providing any piece of music, resulting in 1398 pieces of music being retained. Second, we removed entries which were not changed from the default answers provided in the questionnaire. Third, we processed the onsets of MECs, notably by converting time ranges to individual onsets, by removing some extraneous qualitative comments about specific musical characteristics leading to MECs, and by discarding a few entries which were higher than total track duration. Finally, we cleaned the URLs by removing further qualitative comments and by discarding non-valid URLs. This process resulted in retaining 1187 out of the 1398 reports of pieces of music causing MECs, corresponding to 1150 unique pieces of music associated with 2028 onsets of MECs.

Out of these, 1019 unique pieces of music could be retrieved for the present study, corresponding to 1806 onsets of MECs. We have made the data for these pieces of music available in *Onsets of Chills in Music (oChiM)*, a dataset hosted permanently on the Open Science Framework.[1]

[1] oChiM is available at https://doi.org/10.17605/osf.io/x59fm

Stimulus preparation

In order to be suitable for auditory feature extraction, tracks were retrieved as WAV files, downmixed to mono, and downsampled to 44.1 kHz when necessary, using the *tuneR* R package (Ligges et al., 2018). RMS normalisation was then applied to all tracks simultaneously, using the *soundgen* R package (Anikin, 2019). This process consisted of rescaling all audio files so that they had the same peak amplitude, set at 0 dB, and rescaling them linearly once more so that the RMS amplitude of each file matched that of the file with the lowest RMS amplitude.

5.2.2 Feature extraction

Based on the acoustic and musical elicitors of MECs reviewed above, we extracted a range of features which we believed most closely captured notions of loudness, roughness, brightness, spectral centroid and flux, event density, frequency range, crescendi, tonality, harmony, texture, expectation, and the presence of vocals.

MiningSuite

Most features were extracted using *MiningSuite* (Lartillot, 2019), a MATLAB framework for the analysis of audio and music recordings, which expands on the methods provided by the commonly used *MIRtoolbox* (Lartillot et al., 2008). Default settings for frame size were used for these features, as detailed below.

Loudness and crescendi were approximated with the *aud.envelope* function, which consists of a generic envelope extraction method, further processed following a model of human auditory perception (Klapuri et al., 2006). The feature was extracted with a frame size of 10 ms, corresponding to a sampling rate of 100 Hz.

Spectral frame decomposition was then applied using the *sig.spectrum* function, which applies a Fast Fourier Transform to the audio waveforms of each file, using a sliding window size of 50 ms with 25 ms overlap. This spectral decomposition was used to approximate the following elicitors: roughness with the *aud.roughness* and *sig.flatness* features, respectively estimating sensory dissonance and spectral smoothness; brightness with the *aud.brightness* and *sig.centroid* features, the former capturing the amount of high-frequency energy in the signal and the latter more broadly capturing spectral centroid (the centre of mass of the spectrum—another elicitor of interest); spectral flux with the *sig.flux* feature, calculating the spectral distance between successive frames; event density with the *sig.entropy* feature, computing the relative Shannon entropy of the input (Shannon, 1948); and frequency range with the *sig.spread* feature, capturing variance in the spectrum. These features were all extracted using the same sliding window size of 50 ms with 25 ms overlap as the initial spectral decomposition.

Tonality was approximated with the *mus.key* feature, which estimates tonal centre positions by choosing the highest key candidate from a key strength curve, itself computed by correlating the chromagram of the signal with known key profiles (Gómez, 2006; Krumhansl, 1990). This feature used a sliding window size of 1 s with 0.5 s overlap. Finally, to approximate harmonic change, a six-dimensional tonal centroid was first extracted using the *mus.tonalcentroid* function, corresponding to chord projections on the circle of fifths, minor thirds, and major thirds, before being processed with the *mus.hcdf* harmonic change detection function (HCDF), which returns the flux of the tonal centroid, using the default settings of a 743 ms sliding window size with 74.3 ms overlap.

Due to the computationally intensive nature of extracting many features from many tracks, the scripts were run in parallel on a series of Linux-based compute servers provided by Queen Mary University of London.

Spleeter

To generate a continuous, binary feature representing the presence of vocals, tracks were first processed using the *Spleeter* source separation library (Hennequin et al., 2020), written in Python. The library provides pre-trained models to perform source separation of a music track into two, four, or five stems containing separate instruments. For the present study, two-stem separation was conducted, resulting in two separate tracks containing vocals and accompaniment for each track.

We applied an amplitude threshold to the tracks containing vocals only, in order to generate a binary feature reflecting the absence or presence of vocals which would discard the small amount of residual noise left in the tracks. In practice, since amplitude is characterised by a high degree of zero-crossing, which is not suitable for thresholding, we first computed the loudness of the vocals from their amplitude using the *soundgen* R package, which provides a function allowing the estimation of subjective loudness in sones (a psychoacoustic unit of perceived loudness) for each 20 ms sliding window with 50% overlap, resulting in a value capturing subjective loudness every 10 ms. To prevent the application of a loudness threshold from returning an overly sensitive vocals detection feature, we then applied a rolling maximum filter with a span of 510 ms for each track, before finally applying a loudness threshold, categorising vocals as present if above 2.5 sones, and absent if not. This thresholding process is visualised in Figure 5.1, and resulted in a continuous, binary feature representing the presence

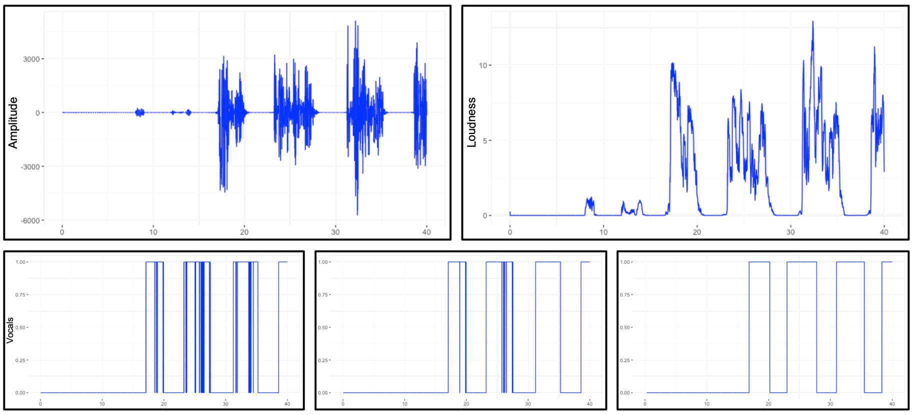

Figure 5.1: Vocals thresholding process. Data is shown for a short, 40-second excerpt in order to illustrate the thresholding process used to identify the presence of vocals. The top row shows the amplitude of the wave form of the vocals track extracted using Spleeter on the left, and an estimate of subjective loudness on the right. The bottom row shows the results of applying a loudness threshold at 2.5 sones with no rolling maximum filter on the left, a rolling maximum filter with a 50 ms span in the middle, and the same filter with a 510 ms span on the right. Several combinations of values were tested before the values used in the last plot were retained, as they were deemed to most closely approximate what manual annotation would return.

of vocals with a frame size of 10 ms.

Note that both the 510 ms value for the span of the maximum filter and the 2.5 sones value for the loudness threshold were chosen manually, after comparing the outputs when changing these two parameters. A systematic validation was not possible due to the absence of labelled data in our sample. However, we conducted several manual checks on a representative set of recordings, and deemed the feature resulting from these parameters as close as possible to what would have resulted from manual annotation of the tracks. This allowed us not to allocate a disproportionate amount of time to the extraction of a single feature in the wider context of the present analysis. Note also that, while this feature is sensitive to differences in loudness between tracks, such concerns were mitigated by the fact that RMS normalisation was conducted prior to this step of the analysis.

IDyOM

Finally, in order to extract information about melodic expectation, it was necessary to extract melodies in MIDI format from each audio track. First, the time-series of continuous frequency values in Hz was extracted for each melody using the *MELODIA* plugin (Salamon & Gómez, 2012) for *Vamp*,[2] via its associated *vamp* Python library.[3] MELODIA enables the estimation of the fundamental frequency of the pitch of the primary melody in polyphonic tracks, and was therefore particularly well-suited for this task.

Then, a simple heuristic provided by the author of MELODIA[4] was re-implemented in R in order to quantise pitch frequencies into discrete MIDI notes. This heuristic consisted of converting each value in Hz to its closest MIDI note before applying a median filter with a 250 ms span in order to remove some of the noise in the underlying data, and finally discarding notes shorter than 100 ms in duration. The resulting MIDI notes were stored in a text file suitable for the next step of the feature extraction process, while note onsets were stored in a reference file separately to convert the extracted features back to time-series suitable for model training.

IDyOM, standing for Information Dynamics of Music (Pearce, 2005, 2018), was used to extract information about melodic expectation from these sequences of MIDI events. IDyOM is a system based on variable-order Markov models, which learns from the statistical regularities in symbolic, sequential events such as MIDI representations of melodies, and applies that knowledge to estimate the likelihood of each event within a sequence. This takes the form of two distinct measures, among other outputs of the model: *entropy*, a measure of uncertainty about which event is predicted to come next given the current context at a specific position in a sequence, and *information content*, a measure of the amount of information that is provided by an event given its previous context, which can therefore be used to quantify the surprisal of the event. For instance, if at time $t - 1$ in a melody, the system is very certain about which note comes next, this will be reflected by low entropy for that note occurring at time t. If the actual note in the sequence was indeed very predictable, this will be reflected by low information content, but if it was unexpected instead, information content will be high.

IDyOM provides a multiple-viewpoint system, which allows sequences to be modelled based on a range of melodic and rhythmic properties, such as pitch, chromatic pitch interval, contour, onset, duration, inter-onset interval, and many more. Options are provided for model training, including training a separate model on each sequence to make predictions for that sequence only (short-term model, capturing dynamic expectation), pre-training a model on a given corpus of sequences (long-term model, capturing schematic expectation), or combining both of these approaches. IDyOM has been the subject of substantial empirical

[2]https://vamp-plugins.org
[3]https://pypi.org/project/vamp
[4]https://github.com/justinsalamon/audio_to_midi_melodia

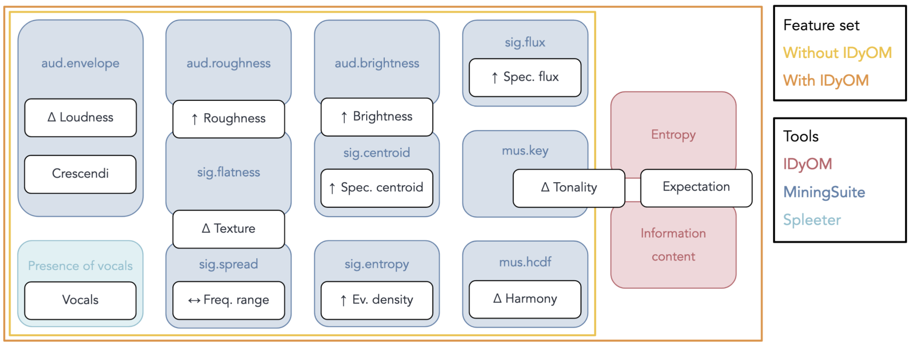

Figure 5.2: List of features extracted from each track. The features are shown in rectangular boxes with coloured backgrounds, corresponding to which tool was used to extract them. The acoustic and musical elicitors of MECs they aim to characterise are shown in white rectangular boxes (using the abbreviations *freq.* for "frequency", *spec.* for "spectral", and *ev.* for "event"), along with the hypothesised direction of their respective effects, as identified from prior research and shown with the following symbols: Δ for changes, $\uparrow$ for increases or elevated levels, and $\leftrightarrow$ for expansions. The outer rectangles represent the two feature sets used for model training, as described later in the present chapter.

testing, and was found to accurately predict melodic expectation in a range of experiments (for an extensive review, see Pearce, 2018).

In the present study, the *cpint* viewpoint (chromatic pitch interval) was derived from the *cpitch* viewpoint (MIDI pitch number), in order to capture the entropy and information content of each MIDI event based on chromatic pitch interval. The model used a combination of the short-term and long-term models, with the long-term model being trained on a set of folk ballads from Nova Scotia, Bach choral melodies, and German folk songs from the Essen Folk Song Collection (for a description of this corpus, see Pearce, 2005). The resulting entropy and information content values associated with each MIDI event for each tracks were linked back to the note onsets previously kept aside in a reference file, allowing these values to be converted to continuous features synchronised with the tracks.

The whole feature extraction process is visualised in Figure 5.2, along with the source of the features and the elicitors they were intended to approximate.

5.2.3 Feature preparation

Key distance

Most features required additional processing in order to capture the hypothesised elicitors laid out in prior research. Tonality, notably, is only thought to affect the occurrence of MECs in occasional cases of changes in tonality. However, the current feature extracted with *mus.key* only captured the tonal centre at a given time, which should have no bearing on the occurrence of MECs. We decided to compute key distance from this feature, following the intuition that more unexpected changes in tonality might be more conducive to experiencing MECs. To do so, we first smoothed the feature by applying majority voting with a 3.5 s span, before assigning a value to tonality changes based on distance on the circle of fifths. For instance, if the tonal centre at time $t-1$ was C, the key distance at time t would be 0 if the tonal centre remained C, 2 if it changed to D, or a maximum of 6 if it changed to F#. The feature was then upsampled to 100 Hz for consistency with the other features, replacing the newly introduced missing values by their nearest existing key distance value.

Interpolation and smoothing

Cubic spline interpolation was applied to all the other auditory features extracted with MiningSuite in order to match this 100 Hz frame rate, with the exception of envelope, which was already sampled at 100 Hz. Care was taken only to allow the addition of new values in the gaps between two existing values, in order to prevent cubic spline interpolation from returning wildly unlikely values at the very beginning and end of each track, where more missing values might be found. Following this, all features (including envelope) were smoothed by using a median filter with a span of 50 ms.

First-order difference

As discussed earlier (and seen in Figure 5.2), many elicitors of MECs refer to changes in acoustic and musical properties, as opposed to specific values. To allow the analyses in the present study to detect this behaviour, first-order differences were computed for each feature and also included in the analyses, such that if a feature had a value of 3 at time $t-1$ and a value of 7 at time t, its first-order difference at time t would be $7-3=4$.

Segmentation

In order to speed up computations for the planned analyses, and to investigate both central tendencies and variance for each feature, summary statistics were computed over successive segments for each feature. All analyses in the present study were run twice: once with a segment size of 200 ms, and once with a segment size of 500 ms, since there was no information to determine a priori which segment size would work best to investigate the occurrence of MECs. For each segment, the mean and standard deviations were computed, resulting in four dimensions for each original feature: μ_0 and σ_0, the mean and standard deviation of the original values, and μ_1 and σ_1, the mean and standard deviation of the first-order difference. Note that due to the way key distance was generated, standard deviations were not computed as they would not contain any meaningful information. In the present study, μ_0 can be thought of as the original feature, σ_0 as the variance in that feature, μ_1 as the rate of change in the feature, and σ_1 as the variance in that rate of change.

In addition, we predicted that for three features (envelope, HCDF, and key distance), changes on a slower time scale might better capture their hypothesised role as elicitors of MECs. These features were therefore also segmented using a 2 s sliding window with 50% overlap, and upsampled to both 200 ms and 500 ms frame sizes, in order to be included in both sets of analyses. The full list of features is shown in Table 5.1. These features were used for both sets of analyses—one with features using a 200 ms frame size, and one with features using a 500 ms frame size, as a result of the two different types of segmentation. The table includes display names for each feature, which are used for plotting the results in the rest of the present chapter due to space constraints within the plots. The computed features have also been made available on oChiM.[5]

Table 5.1: Display names for features included in both sets of analyses

Tool	**Feature category**	**Features**			
		μ_0	σ_0	μ_1	σ_1
MiningSuite	Envelope	$\mu_0$env	$\sigma_0$env	$\mu_1$env	$\sigma_1$env
	Envelope +	$\mu_0$env+	$\sigma_0$env+	$\mu_1$env+	$\sigma_1$env+
	Roughness	$\mu_0$rough	$\sigma_0$rough	$\mu_1$rough	$\sigma_1$rough
	Flatness	$\mu_0$flat	$\sigma_0$flat	$\mu_1$flat	$\sigma_1$flat
	Brightness	$\mu_0$bright	$\sigma_0$bright	$\mu_1$bright	$\sigma_1$bright
	Spec. centroid	$\mu_0$spcent	$\sigma_0$cent	$\mu_1$cent	$\sigma_1$cent
	Spec. flux	$\mu_0$spflux	$\sigma_0$flux	$\mu_1$flux	$\sigma_1$flux
	Spec. entropy	$\mu_0$spent	$\sigma_0$spent	$\mu_1$spent	$\sigma_1$spent
	Spec. spread	$\mu_0$spspr	$\sigma_0$spspr	$\mu_1$spspr	$\sigma_1$spspr
	Key distance	$\mu_0$kdist		$\mu_1$kdist	
	Key distance +	$\mu_0$kdist+	$\sigma_0$kdist+	$\mu_1$kdist+	$\sigma_1$kdist+
	HCDF	$\mu_0$hcdf	$\sigma_0$hcdf	$\mu_1$hcdf	$\sigma_1$hcdf
	HCDF +	$\mu_0$hcdf+	$\sigma_0$hcdf+	$\mu_1$hcdf+	$\sigma_1$hcdf+
Spleeter	Vocals	$\mu_0$voc	$\sigma_0$voc	$\mu_1$voc	$\sigma_1$voc
IDyOM	Mel. entropy	$\mu_0$melent	$\sigma_0$melent	$\mu_1$melent	$\sigma_1$melent
	Mel. IC	$\mu_0$melic	$\sigma_0$melic	$\mu_1$melic	$\sigma_1$melic

Note. Four types of features for each feature category. μ_0 = mean of the original values, σ_0 = standard deviation of the original values, μ_1 = mean of the first-order difference, σ_1 = standard deviation of the first order difference, + = feature segmented using a longer, sliding window, Spec. = Spectral, Mel. = Melodic, IC = information content.

[5]https://doi.org/10.17605/osf.io/x59fm

5.3 Analysis

5.3.1 Permutation tests

In order to identify patterns in the behaviour of each feature around the onset of MECs, we ran a series of permutation tests (for a similar approach, see Grewe et al., 2009b). As previously discussed in Chapter 3, permutation tests consist of identifying a test statistic that captures the dimension we want to measure, and generating a null distribution by permuting samples, in order to assess whether or not the observed results were unlikely enough to reject the null hypothesis. Here, we explored each feature by evaluating how unlikely their values were around the onset of MECs, when compared to other moments within the tracks which were not reported as causing MECs.

To do so, we extracted all 20-second excerpts centred around the onsets of MECs from each track. We only retained complete excerpts, meaning that onsets of MECs within the first or last 10 seconds of each track were discarded. Rather than choosing individual control excerpts for each excerpt causing MECs, we split every track into sequential 20-second excerpts, discarded every excerpt which was not complete, or partially or fully overlapped with any of the excerpts causing MECs, and retained all the remaining excerpts as controls. This process therefore resulted in two unbalanced sets of 20-second excerpts, capturing 10 seconds before and after each onset of MECs for one set, and almost all other moments within the tracks for the control set.

The test statistic was computed for each frame of each feature, and consisted of the difference between the average values for excerpts causing MECs and control excerpts. Two-tailed permutation tests were run by randomly permuting the excerpts while keeping the same number of excerpts in each set, using Monte Carlo estimation with 5000 replications. We only ran the permutation tests on the data with a 500 ms frame size in order to limit the number of comparisons we would draw, and we used Bonferroni correction within each feature, to further mitigate the fact that, even with a 500 ms frame size, 41 significance tests would be required for each feature.

5.3.2 Principal components analysis

We trained models to assess whether or not the onsets of MECs could be predicted using audio-derived acoustic and musical features, and if so, which features were most important in driving such predictions. In addition to the two types of segmentations discussed above, we evaluated two types of models (described later in this section), on two sets of features. The first set of features did not include the IDyOM features, while the second set did. This was done in order

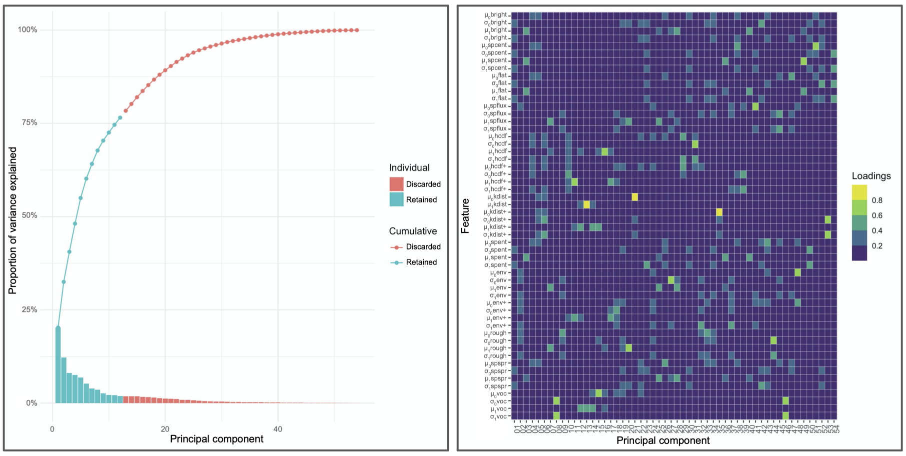

Figure 5.3: Visualisation of principal component analysis for one of the four combinations of feature set (without IDyOM features) and frame size (200 ms). On the left, a scree plot shows the 12 retained principal components, along with the proportion of variance explained they account for, both individually and cumulatively. On the right, a heat map displays feature loadings on each principal component.

to specifically evaluate the effect of accounting for expectation on the predictive performance of the models. For clarity, we hereafter refer to these differences in model training as differences in *frame size* (200 ms or 500 ms), *feature set* (without or with IDyOM features), and *model type* (described below).

We expected a high degree of collinearity in the features, with features in some cases being exactly identical, with the exception of the type of segmentation they were subjected to. As discussed in Chapter 4, PCA is a convenient way to address collinearity, by retaining principal components above a set eigenvalue threshold. We therefore conducted four PCAs with centring and scaling—one for each combination of frame size (200 ms or 500 ms) and feature set (with or without IDyOM).

The results of the PCAs are discussed here for simplicity, as they are not crucial to the rest of the findings discussed in the present chapter. We retained 12 principal components with eigenvalues above one for each PCA using the first feature set (without IDyOM) regardless of frame size, and 15 principal components for each PCA using the second feature set (with IDyOM) regardless of frame size as well. Taken together, these principal components accounted for at least 75% of cumulative proportion of variance explained for each PCA (see Figure 5.3 for an example).

While we didn't attempt to interpret how the features were grouped together into specific principal components, we stored feature loadings on each principal component and proportion of variance explained by each principal component for later analyses of feature importance.

5.3.3 Hidden Markov models

The first models we trained were hidden Markov models (HMMs). As opposed to Markov chains, which model the probabilities of sequences of observable states, HMMs model the probabilities of hidden states, which themselves drive the probability distributions of observable events. HMMs are specified by a set of of hidden states, the transition probabilities between these states, an initial probability distribution for these states, a set of observations, and the observation likelihoods associated with each state, also called emission probabilities——the probability that an observation was generated by a specific state (for excellent introductions to HMMs, see Jurafsky & Martin, 2021; Rabiner, 1989).

In its most simple form, for a univariate sequence of observations, an HMM associates each hidden state with a specific emission distribution (e.g., normal distribution with a given mean and standard deviation) of the observed values in that sequence. In practice, we are often faced with multivariate sequences of observations, and instead of using multivariate Gaussian emissions, which might be limited in how accurately they can represent observations, Gaussian mixture model emissions (GMMs) are often used. GMM-HMMs are particularly well suited to modelling auditory events due to their flexibility and sequential nature, and have successfully been applied to speech recognition (see Juang & Rabiner, 1991), or, in a music context, to segmentation, genre classification, sequence prediction, and event detection (e.g., Ajmera et al., 2003; Wang et al., 2019).

It is worth clarifying what each component of an HMM represented in the present analysis. The observations corresponded to the multidimensional sequence of features (or rather, of principal components), different configurations of this multidimensional sequence were modelled by GMMs, each governed by a different hidden state of the HMM. It was assumed that specific sequences of hidden states gave rise to the occurrence of MECs (as opposed to having a single hidden state represent MECs specifically).

Fundamentally, HMMs are characterised by three problems (Jurafsky & Martin, 2021; Rabiner, 1989): the likelihood problem (determining the likelihood of a specific sequence of observations given the HMM and observations), the decoding problem (discovering the best hidden state sequence given the HMM and observations—not relevant to the present study because we did not attempt to interpret the hidden states themselves), and the learning problem (learning the HMM parameters given the states and observations). The present automatic MEC onset detection task consisted of two steps. In the first step, we trained two separate HMMs (learning problem). The first HMM was trained on excerpts centred around the onsets of MECs, and the second HMM on control excerpts. In the second step, we presented these trained HMMs with new excerpts to obtain the likelihood of each excerpt according to each HMM (likelihood problem). If the HMM trained on excerpts causing MECs returned the highest likelihood, the excerpt was categorised as inducing MECs, and vice versa. Implementation details are provided below.

As with the permutation tests, this approach required extracting two sets of excerpts, since HMMs are best trained on short sequences of equal length. The exact same procedure was used, selecting excerpts centred around the onset of MECs, and control excerpts sequentially through the rest of each track. However, instead of 20-second excerpts, we tested two excerpt durations: two seconds (with a frame size of 200 ms only), and five seconds (with a frame size of 200 ms or 500 ms), in order to speed up computations while ensuring a reasonable amount of frames would be retained for each excerpt. *Excerpt size* (2 s or 5 s) therefore corresponds to the last aspect of model training we controlled, in addition to frame size, feature set, and model type.

In the survey from which onsets of MECs were gathered, participants reported these onsets with a time resolution of one second. We therefore decided to augment the training data by also extracting excerpts categorised as causing MECs for each frame within one second of the original onset of MECs. For instance, with a frame size of 500 ms, instead of getting a single excerpt for an onset of MECs at $t = 40$ s, we extracted five excerpts at $t = 39$, 39.5, 40, 40.5, and 41 s. This also allowed us to reduce the severe class imbalance, with control excerpts outnumbering excerpts causing MECs by a ratio of 100:1.

Using the *pomegranate* Python library (Schreiber, 2018), a pair of HMMs (MECs and control) was trained for each combination of feature set (with or without IDyOM), frame size (200 ms or 500 ms), and excerpt size (2 s or 5 s), using manual grid search by iterating over the number of hidden states and the number of Gaussian mixtures the models should be trained with. More specifically, in order to save time on this very computationally intensive training process, we trained each model using odd numbers of states and mixtures (ranging from 1 to 17), before testing the even numbers of states and mixtures nearest to the best performing model. This process is illustrated in Figure 5.4. We used five-fold cross-validation, using for each fold 60% of the tracks as a training set, 20% as a validation set to pick the best performing model for testing, and 20% as a testing set to return final performance metrics for the combination of learning parameters which performed best across all five folds. As with feature extraction, model training was run in parallel on university-provided compute servers.

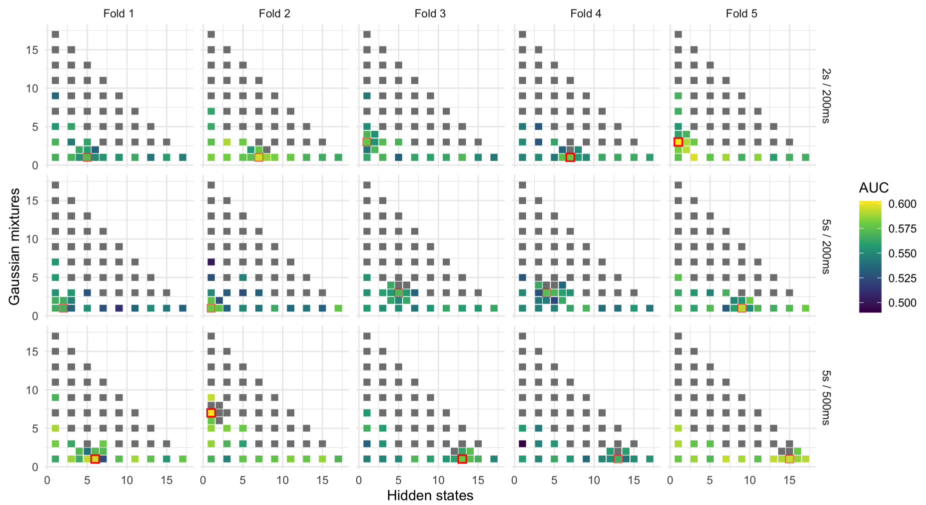

Figure 5.4: Grid search for GMM-HMM training. In this example, HMMs were trained on the feature set which includes IDyOM features. A pair of HMMs (MECs and control) was trained for each combination of number of hidden states (bottom axis), Gaussian mixtures (left axis), cross-validation fold (top axis), and excerpt size and frame size (right axis). Model performance was evaluated using area under the receiver operating characteristic curve (AUC). Odd numbers of states and mixtures were first used for training. For each combination of fold, excerpt size, and frame size, additional models were trained using the even numbers of states and mixtures closest to the best performing models (highlighted in red), resulting in some cases in increased performance (e.g., see Fold 1, 5 s excerpt size, 500 ms frame size). Cell colour represents models which failed to converge, in grey, or AUC. The best performing HMMs for each fold are highlighted with a thicker, red border.

The process for training an individual HMM was as follows. All data was concatenated, ignoring its sequential nature, to identify a cluster for each hidden state using k-means, and initialise the parameters of the corresponding GMM using said cluster. The model was then initialised with a uniform probability transition matrix, before training began using the Baum-Welch algorithm (see Jurafsky & Martin, 2021). Regularisation was applied, by setting a transition and emission pseudocount of 0.1, and an edge and distribution inertia of 0.1 (see Schreiber, 2018). If the model failed to converge, generally due to underflow errors, training was attempted once more. If unsuccessful, training was abandoned and a new model was trained for the next step of the grid search, as seen in Figure 5.4.

Full tracks, split into consecutive excerpts, were used for model validation. For each excerpt from each track in the validation test, a log probability value was obtained for each of the pair of trained HMMs (the one trained on excerpts causing MECs, and the one trained on control excerpts), using the forward algorithm (see Jurafsky & Martin, 2021). If the HMM trained on excerpts causing MECs returned the highest log probability for the tested excerpt, that excerpt was predicted as an occurrence of MEC, and vice versa. For each fold, this validation process resulted in a univariate time-series of binary predictions for each track.

Due to the sequential nature of the resulting predictions, two sets of performance metrics were computed, using the *sed_eval* Python library (Mesaros et al., 2016): *event*-based metrics, and *segment*-based metrics. Event-based metrics compare ground truth and predictions frame by frame, usually allowing for a *collar*—a small amount of tolerance around the onset of a predicted occurrence of MECs. We opted for a one-second collar, which resulted, for instance, in predicted occurrences of MECs being categorised as true positives if they were within one second of an onset of MECs in the ground truth. Segment-based metrics compare ground truth and predictions in a fixed time grid, marking segments are active if they include an onset of MECs, and inactive if not. We opted for a five-second segment length, resulting in a given segment being categorised as a true positive if it included an onset of MECs in both the ground truth and the predictions.

For both types of metrics, we computed precision, recall, and F-measure—the harmonic mean of precision and recall. In addition, for segment-based metrics, we computed balanced accuracy, which is not available for event-based metrics (see Mesaros et al., 2016), as well as true positive rate and false positive rate, in order to compute the area under the receiver operating characteristic curve (AUC). In a typical classification task, the AUC is calculated by modifying the classification threshold. In the present analysis, however, predictions were not made based on a classification threshold, but rather by comparing log probabilities between two HMMs. To emulate the principle of a classification threshold, we collected all frame-wise log probability differences and extracted their percentiles. Each percentile was used as a proxy for a threshold, by generating a new set of predictions based on whether or not the difference between the log probabilities of both HMMs was higher or lower than that percentile. This allowed us to collect 100 pairs of true positive rates and false positive rates (one for each percentile), which were then used to compute the AUC with the *scikit-learn* Python library (Pedregosa et al., 2011).

After observing the results, discussed later in the present chapter, we also decided to compute F_β, which applies an additional weight β to the F-measure in order to disproportionately favour precision or recall over the other. We picked a value of 2 for β—a standard value to signify that we considered recall twice as important as precision in the present analysis. This decision and its implications are explored in the discussion. For the validation sets, only AUC was used to select the combination of learning parameters which performed best across all five folds. For each fold, the HMMs trained using these learning parameters were then used on the testing set to compute the AUC. All other performance metrics were computed for the classification threshold (i.e., the log probability difference threshold) which returned the highest F_β value. These metrics were then averaged across all five folds to return final model performance metrics.

5.3.4 Support-vector machines

For purposes of comparison, we ran all analyses using a second type of model called support-vector machine (SVM). For classification tasks using an SVM, model training consists of finding the hyperplane which maximises the distance between the two classes, and is used as a decision boundary to make predictions. SVMs are widely used for classification tasks, and were chosen in the present analysis as a more naive modelling approach because, when compared to HMMs, they are much quicker to train, less prone to fitting errors, relatively easy to interpret, and have the added benefit of being more forgiving in terms of statistical assumptions than logistic regression—another commonly used classification method. However, as opposed to HMMs, they do not take into account the sequential nature of the data.

SVMs were trained using *scikit-learn* with a linear kernel, no random feature selection, a stopping criteria set at 10^{-5}, and by adjusting weights inversely proportional to class frequencies in order to account for class imbalance. The HMM workflow detailed above was replicated to train, validate, and test SVMs, with a few exceptions. First, separating the training data into excerpts was not necessary, since SVMs could be trained on the whole dataset at once. This also meant that instead of generating several excerpts to account for the imprecision in the time resolution of onsets of MECs, we simply assigned positive labels to all frames within one second of the onset of MECs. Second, grid search was performed, but only involved varying feature set (with or without IDyOM), frame size (200 ms or 500 ms), and the regularisation parameter for linear SVMs (using a logarithmic scale ranging from 10^{-24} to 10^{4}, with even-numbered exponents only). Third, predictions were made frame-wise, but we still used event-based and segment-based evaluation metrics. Finally, as discussed earlier, computing the AUC requires modifying the classification threshold, but linear SVMs do not return the probabilities associated with the predictions for each frame by default. To obtain these probabilities, we applied Platt scaling (Platt, 1999) on the decision function, and then computed the AUC similarly as for the HMMs, by adjusting the classification threshold over the percentiles of these probabilities.

5.3.5 Feature importance

The final step of the analysis was to extract feature importance. To simplify the interpretation of the results, feature importance was only extracted from the best performing model across all combinations of excerpt size, frame size, feature set, and model type. Since the best performing model ended up being an SVM, as revealed in the results, the method described below is specific to feature importance extraction from the parameters of a trained SVM.

Using Platt scaling involves a second layer of cross-validation within each cross-validation fold. We refer to the folds from this second layer as *SVM folds* to differentiate them from the cross-validation folds used in the workflow detailed above. First, feature coefficients were extracted from the parameters of the trained models. Since we were interested in assessing feature importance, as opposed to trying to interpret the directionality of the coefficients (which was more relevant to the permutation test analyses), absolute values were taken, therefore preventing coefficients from cancelling each other out when averaged over SVM folds, if directionally different. These absolute values for the coefficients of each feature were averaged over each SVM fold, resulting in a set of five positive, averaged coefficients for each feature, i.e., one coefficient per cross-validation fold.

However, in this case, the features were principal components obtained with PCA. To tie the coefficients back to the original features, we had to apply weights to these coefficients based on proportion of variance explained by and feature loadings on each principal component, thereby assigning a share of the influence of each principal component on the models to its constituent features, based on how much variance in the data that principal component accounted for. To do so, we simply took the feature loadings on each principal component, multiplied them by the proportion of the variance explained by each principal component, and multiplied that by the coefficients extracted from the models for that principal component. We then added up the coefficients for each feature across all principal components, resulting, again, in a single coefficient per feature for each of the cross-validation folds (but this time, for each of the original features, as opposed to principal components).

The magnitude of the coefficients differed between cross-validation folds, but since the folds were of equal size, and we were interested in feature importance, we rescaled the coefficients linearly for each fold, such that 0 corresponded to the least contributing feature for that fold, and 1 to the most contributing feature. Finally, we averaged these values across the five cross-validation folds, resulting in a single feature importance value, between 0 and 1, for each of the original features that were used for PCA before model training. Note that these values

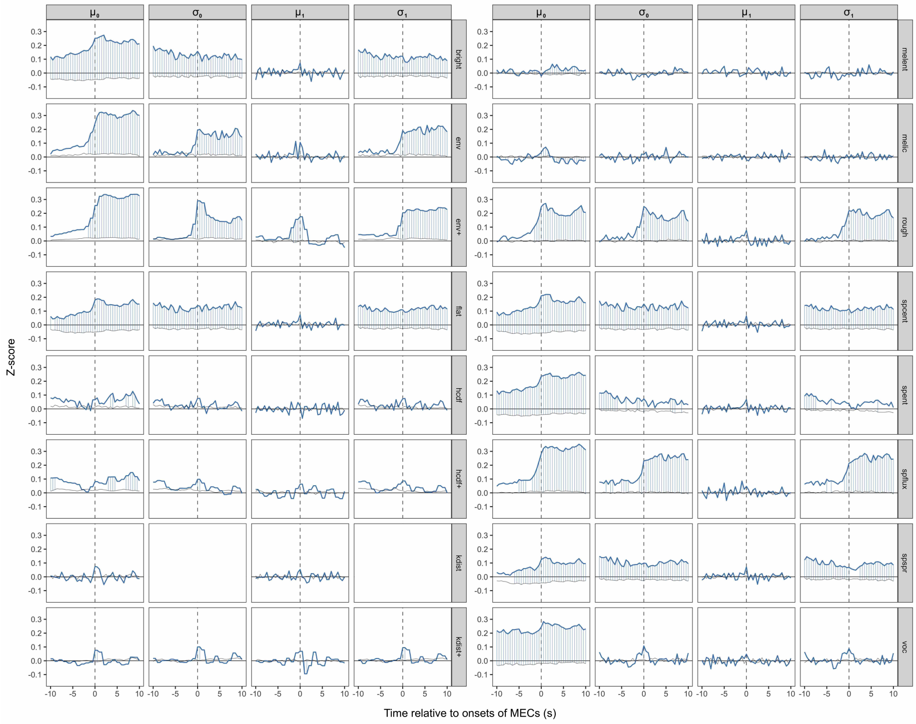

Figure 5.5: Visualisation of the permutation tests. Vertical blue lines denote frames for which there were significant differences between the excerpts causing MECs (thick blue line) and control excerpts (thin grey line).

do not exactly correspond to a rank, but rather, to a continuous spectrum of feature importance, where a value of 0 would represent the feature with the lowest feature importance on all five cross-validation folds, and a value of 1 for the highest feature importance.

5.4 Results

Far too many permutation tests were conducted to meaningfully present the results in a table. Instead, we opted to visualise all the results in Figure 5.5. This figure warrants extensive explanation, and is therefore discussed here rather than in the figure caption. Note that all feature values were Z-scored in the figure, in order to better visualise the magnitude of the effects, and to enable better comparisons across features.

Each plot within the grid corresponds to one of the features that was included in the PCA (keeping in mind that we only ran permutation tests on the dataset with a 500 ms frame size). The plots are arranged by rows, corresponding to feature categories, and by columns, corresponding to summary statistics. The x-axes represent 20-second excerpts, centred around the onsets of MECs, and the y-axes represent Z-scores. Within each plot, the horizontal line at $y = 0$ therefore represents a Z-score of 0, and the vertical dotted line at $x = 0$ represents the onsets of MECs.

The thin grey time-series represent values averaged over all the control excerpts, while the thicker blue time-series represent values averaged over all the excerpts centred around reported onsets on MECs. For a given frame, a significant difference in average feature values between these two time-series, as revealed by a permutation test, is indicated by a vertical blue line. As mentioned earlier, permutation tests were not conducted for $\sigma_0 kdist$ and $\sigma_1 kdist$ due to the way these features were computed.

Interpretation of these results is provided in the discussion, but we can already easily identify several features showing higher values in the excerpts causing MECs, or sharp increases around the onset of MECs.

The results for the predictive modelling of the onsets of MECs are presented in Table 5.2, which shows evaluation metrics for the best performing combinations of model type (HMM or SVM), feature set (with or without IDyOM features), and metrics type (segment-based or frame-based). Including IDyOM features in the feature set resulted in slightly better metrics overall, with the exception of AUC for the HMM, which was slightly lower with IDyOM features. Both SVMs outperformed the HMMs, with the SVM trained using all features reaching a segment-based AUC of 0.597, F_β of 0.167, and balanced accuracy (the arithmetic mean of the accuracy on each class) of 0.580. All models exhibited very low precision and good recall, therefore leading to low F-measures and motivating the choice of F_β as an evaluation metric, as expanded upon in the discussion. Interestingly, the best performing models for each category, as described in Table 5.2, all used the features with a 500 ms frame size, as opposed to 200 ms.

Table 5.2: Evaluation metrics and learning parameters for each model

Model	**IDyOM**	**Metric type**	**Evaluation metrics**					
			AUC	F_β	F	P	R	BA
HMM	✗	Segment	0.579	0.161	0.075	0.040	0.682	0.564
		Frame	-	0.091	0.042	0.022	0.442	-
	✓	Segment	0.577	0.165	0.076	0.040	**0.744**	0.571
		Frame	-	0.099	0.046	0.024	0.429	-
SVM	✗	Segment	0.592	0.166	0.078	0.041	0.713	0.579
		Frame	-	0.080	0.036	0.019	0.466	-
	✓	Segment	**0.597**	**0.167**	**0.078**	**0.041**	0.693	**0.580**
		Frame	-	0.088	0.041	0.022	0.383	-

Note. Highest values in bold. HMM = Hidden Markov model, SVM = Support-vector machine, AUC = Area under the receiver operating characteristic curve, F_β = F_β-measure, F = F-measure, P = Precision, R = Recall, BA = Balanced accuracy.

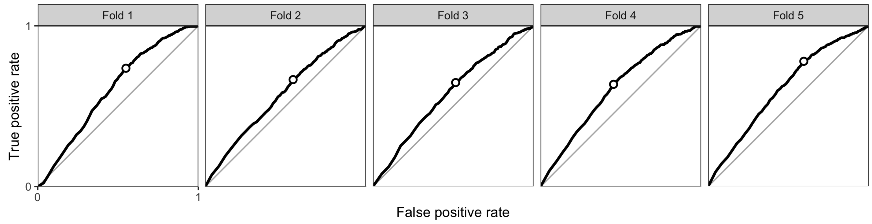

Figure 5.6: Receiver operating curves for each cross-validation fold of the SVM trained using all available features. Overall AUC for the model was computed by averaging the AUCs for each curve. The threshold which returned the highest F_β value is visualised by a circle on each curve, and corresponds to the threshold at which all evaluation metrics were retained before being averaged to return overall evaluation metrics for the model.

In order to maximise the amount of data that could be used for model training, we opted for the cross-validation approach described above, testing the trained model on a different section of the dataset for each cross-validation fold, instead of training a final model to obtain evaluation metrics on a holdout set. Therefore, as explained earlier, AUC was computed from five different receiver operating curves instead of a single one. These curves are shown in Figure 5.6, along with the threshold which resulted in the highest F_β values for each curve.

Finally, the contribution of each feature to the best-performing SVM is visualised in Figure 5.7, in terms of feature importance. Melodic entropy and information content, as computed with IDyOM, were the best predictors of MECs, with melodic entropy reaching a value of 1, meaning it was the most important predictor on all five cross-validation folds. Following these two features were the variance of the first-order difference of melodic entropy, as well as mean spectral flatness, spread, and centroid. The least contributing features were the mean first-order differences of brightness, roughness, and spectral entropy, spread, and centroid.

5.5 Discussion

The permutation tests replicated many of the findings from prior research. Three main types of patterns are identifiable in Figure 5.5: features which were consistently higher than controls around the onset of MECs, features which showed a sharp increase around the onset of MECs, and features which showed no significant differences from controls. First, looking at the original features (as assessed with μ_0), MECs were characterised by elevated brightness, flatness, spectral centroid, spectral entropy, melodic entropy, and presence of vocals, as well as sharp increases in envelope, melodic information content, roughness,

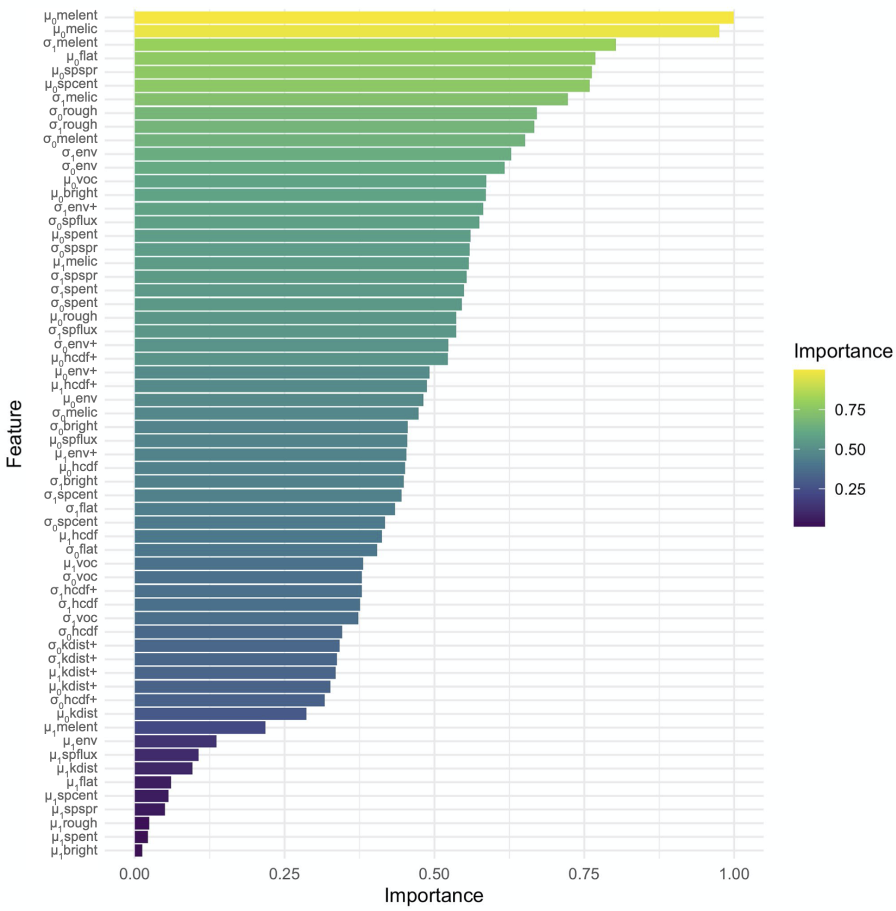

Figure 5.7: Relative feature importance for the SVM trained using all available features. Importance uses an arbitrary scale from 0 to 1, with feature importance rescaled linearly such that the feature contributing the most on each cross-validation fold received a score of 1, as mean melodic entropy did in this case, and the feature contributing the least received a score of 0, which no feature did for all five folds.

spectral flux, spectral spread, and melodic entropy, as well as in envelope and key distance when computed with a larger sliding window. While there were also several frames showing significant differences for the harmonic change detection function (HCDF, and its associated feature computed with a larger sliding window), the overall differences were less convincing than with other features. Interestingly, melodic information content increased around the onset of MECs and decreased afterwards, suggesting the possibility that MECs were associated with a single unexpected event, immediately followed by a return to more expected events. In their study, Cheung et al. (2019) found evidence that information content and entropy interacted in eliciting pleasure, with surprising chords eliciting high pleasure in certain contexts and low pleasure in uncertain contexts. While their experimental paradigm allowed for much more temporal precision than the one presented in this chapter, the very brief decrease in melodic entropy combined with the increase in melodic information content seen in Figure 5.5 around the onset of MECs could support the presence of such an interaction.

The variance in these features (as assessed with σ_0) largely followed similar patterns, with the exception of the presence of vocals, which only contained a single frame with a significant difference between MECs and controls, located at the exact onsets of MECs. We would note that the presence of vocals was the only binary feature in this analysis, and was relatively slow-moving compared to the other features, due to how it was pre-processed. Regardless, this could be interpreted as vocals being more likely around the onsets of MECs (as seen with μ_0), and MECs being slightly but not overly affected by increased variance in the presence of vocals around their onsets. Significant effects were much more sparse when looking at the rate of change of each feature (as assessed with μ_1), with a convincing peak only occurring for the version of envelope computed with a larger sliding window, providing further evidence that sudden, large changes in loudness are associated with MECs. Variance in these rates of change (as assessed with σ_1) almost exactly followed the variance in the original features (as assessed with σ_0), which follows the intuition that when variance in a feature is significantly above average, this is also reflected in the variance of its rate of change.

It is worth pointing out that the patterns seen in each feature were of reasonable magnitude, with most differences ranging from 0.1 to 0.3 in Z-scores. It is also worth noting that these were all hypothesised based on previous research rather than exploratory findings. In addition, we applied extremely strict Bonferroni correction, reducing the alpha threshold for statistical significance to slightly higher than 0.001, and therefore providing a high degree of confidence in the identified effects. Tying these results back to findings from prior research (as listed in detail in the introduction and in Chapter 2), the permutation test analysis provided a large-scale replication of effects showing MECs as being associated with all the acoustic and musical elicitors that were approximated with extracted features, including increases in loudness (e.g., Sloboda, 1991), crescendi (e.g., Panksepp, 1995), increased roughness (e.g., Grewe et al., 2007), brightness (e.g., Bannister & Eerola, 2018), event density (e.g., Nagel et al., 2008), and spectral centroid and flux (e.g., Bannister & Eerola, 2018), expansion of the frequency range (e.g., Guhn et al., 2007), and changes in texture, harmony and tonality (e.g., Sloboda, 1991). Finally, we provided novel quantitative evidence for the existence of effects of vocals and of melodic expectation on the occurrence of MECs, which suggests that MECs are more likely in the presence of vocals, and around the onsets of unpredictable notes in uncertain melodic contexts.

Several models were trained to perform automatic MEC onset detection. This resulted in three sets of findings which warrant further discussion. First, we expected HMMs to perform better than SVMs but the opposite was true. As discussed earlier, HMMs are particularly well suited to auditory event detection tasks. However, in prior research, such events were most often well-defined, with precise, objective onsets. Detecting MECs comes with an additional layer of abstraction, with MECs not being detectable directly from the signal, but rather, being psychophysiological reactions to some properties of the signal. In addition, onsets of MECs were gathered from subjective survey answers and therefore did not represent an exhaustive and precise source of ground truth. It is encouraging, however, that prediction performance was relatively comparable between HMMs and SVMs given these limitations, showing that models such as HMMs can be used to detect the onsets of MECs. It could be that HMM performance could be improved by using longer frame sizes and window sizes, for instance, or by better defining an interpretation of the hidden states for each feature. It remains our intuition that sequential models are best suited to the detection of MECs, but successfully applying such methods might rely on obtaining more exhaustive, better-quality data.

Second, while recall was high, precision was very low for all models. Precision, here, refers to the proportion of predicted MECs which were actually MECs, while recall refers to the proportion of MECs in the ground truth which were predicted as MECs. In other words, the models we trained identified a large proportion of the MECs in the ground truth (high recall) but also predicted far too many MECs where none were recorded in the ground truth (low precision). It is worth mentioning, however, that due to the very high class imbalance in the data, baseline precision was extremely low ($P_B = 0.011$), but all the models still performed better than chance as shown by all AUC values exceeding 0.5. It is because of this low precision that we opted to report results based on the highest F_β value returned by each model, since not doing so disproportionately penalised recall by optimising for very small increases in precision. Optimising for recall also made sense, conceptually, due to the nature of the ground truth data. Indeed, since the ground truth data was acquired experimentally, it is highly likely that, with many more participants in the survey, additional onsets of MECs would have been reported for some of the tracks used for model training, which would have occurred during passages categorised as controls in the present analysis. This made the ground truth incomplete, which motivated attempting to predict all the MECs reported in the data by maximising recall, rather than only the MECs present in the data by maximising precision, with the added advantage that an F_β-optimised model should provide more useful information about which features best characterised the onsets of MECs when looking at feature importance. While these results are not reported in the present chapter for brevity and clarity, it is also worth mentioning that reporting metrics based on the highest F_β value as opposed to the highest F-measure came at almost no cost to precision.

Finally, we found that the models which included IDyOM features in the training set outperformed the models which didn't. While these performance improvements were small, it is worth keeping in mind that, as opposed to the other features which benefited from robust data extraction methods, melodic entropy and information content were both computed based on automatically extracted melodic pitch frequencies—a notoriously difficult and imprecise process subject to much ongoing research—which were then converted to sequences of notes based on a simple heuristic. Despite these limitations, these findings correspond to quantifiable effects of expectation on the occurrence of MECs, which are further reinforced by the findings on feature importance discussed below.

Melodic entropy and information content were indeed the best predictors of MEC onsets by far, followed by the variance in their rates of change as well as the mean of some spectral features (flatness, spread, and centroid). While feature importance is not that informative on its own, it provides interesting insights when combined with the results from the permutation tests, by allowing comparisons between the magnitude of the difference between MECs and controls in each feature, and the degree to which these differences were predictive of MECs. For instance, the lack of detected effects of first-order differences on MECs is also apparent in the fact that most of these features were not strong predictors of MECs. However, these comparisons can be more ambiguous. Many features, such as sharp increases in loudness, benefit from extensive prior empirical support, and strongly displayed the expected behaviour in the permutation tests, but they were not highly ranked in terms of predictive performance. Conversely, the magnitude of the effects detected in the permutation tests for melodic entropy and information content was much lower than for other features, and yet, these expectation-related features were the best predictors of MECs. This suggests that, while previous findings about correlations between various features and MECs were replicated, these features were not always strong predictors of MECs. Taking loudness as an example, it could be that MECs are indeed characterised by sharp increases in loudness, but sharp increases of loudness occur at other moments as well, which makes them inaccurate predictors of MECs. In other words, loudness might be a necessary but non-specific elicitor of MECs. Conversely, small differences in expectation between MECs and controls could be due to small effect sizes, or to increases in entropy and information content only being present in a subset of MEC onsets, but appear to be highly specific to MECs, therefore driving their predictive performance.

Interestingly, Bannister (2020c) found that manipulating loudness and brightness around known onsets of MECs resulted in changes in how many people experienced MECs. Given the results of the feature importance analysis discussed above, a possible interpretation for these causal findings could be that a certain combination of factors is necessary for the occurrence of MECs, as seen by the finding by Bannister that some people reported MECs regardless of loudness and brightness manipulations, but that these manipulations acted upon a loudness and brightness threshold at which different people experience MECs. In other words, it could be that specific levels of loudness and brightness were required for the experience of MECs in some participants, but were not necessarily predictive of such experiences on their own.

It is also worth mentioning a recent article by Mori (2022), which was published after the present experiment concluded, and featured a very similar methodology to ours. In the study described by the author, 54 participants who often experience MECs were asked to listen to a few self-selected pop rock songs which induce MECs or tears while indicating occurrences of MECs or tears with buttons presses. Acoustic features were extracted using MIRtoolbox, and included most of the acoustic features used in the present analysis, as well as a range of rhythm-related features and mel-frequency cepstral coefficients. Excerpts centred around the onsets of MECs, tears, and a comparable number of randomly selected control excerpts were labelled to train a multi-class ridge regression classifier. Instead of using AUC, predictive performance was assessed with permutation tests, conducted by randomly resampling the predicted labels for each frame to generate a null distribution of predictive accuracies. Feature importance was assessed with bootstrapping—a preferable approach to the one we used, partly because it allows using each feature directly without the need for dimensionality reduction, but which was not practical for our purposes due to our prohibitively computationally expensive model training process.

Mori (2022) obtained a classification accuracy of 43% around the onsets of MECs, which was significantly better than chance, as revealed by permutation tests. For our best performing model, AUC was 0.597 and balanced accuracy was 58%, though it is worth pointing out that these numbers are not directly comparable, since Mori (2022) conducted multi-class classification on a balanced dataset, and assessed prediction accuracy with different metrics. In addition, the author identified that these predictions were significantly driven by minor mode immediately preceding MECs (highlighting a possible relationship with expressed valence as seen in Chapter 4), higher event density and rhythmic entropy at the onset of MECs, and higher spectral flux and rhythmic entropy after the onset of MECs (replicating some of our results as well as some hypothesised effects of rhythmic properties detailed in Chapter 2). Various mel-frequency cepstral coefficients also significantly predicted MECs, but these provide less room for interpretation. The author conducted an additional study to investigate the effect of lyrical content, as extracted using natural language processing, but found no effect of lyrics on MECs. Overall, Mori (2022) mainly attributed the occurrence of MECs to violations of rhythmic expectation, which is certainly complementary to the present findings about the effects of melodic expectation on MECs.

Our study suffered from a few limitations, the main one being that a model with an AUC below 0.60 is generally considered to be a very poor classifier. However, it is worth placing model performance within the context of the task at hand. As discussed throughout the present chapter, we expected automatic detection of MEC onsets to be far from trivial, due to the inherent impossibility in collecting exhaustive ground truth data, and to MECs being a subjective, psychophysiological response known to be caused by a wide range of elicitors, some of which would be exceedingly difficult to quantify for modelling purposes. Throughout the modelling process, we had to make a lot of necessary decision with little existing basis for an informed choice between the available options. Given this, we aimed to reach a balance between time constraints and informed guesses as to which processing steps would lead to the best chances of success. These decisions were all documented throughout the present chapter in order to provide transparency and enable reproducibility. One decision, notably, was to automatically extract melodic pitch frequency in order to compute expectation-related features. It is highly likely that a replication of the current study using transcribed melodies would lead to a more precise understanding of the effects of expectation on MECs.

More importantly, we wanted our models to be interpretable, which meant that we needed to use both interpretable features, and interpretable models, at the necessary cost of predictive performance. It is highly likely that more powerful approaches, such as neural networks or ensemble methods, using features such as mel-frequency cepstral coefficients as model inputs, would result in better predictive performance. These approaches might gain from some methodological insights gathered in the present study, such as the use of features computed over large sliding windows and segmented using a 500 ms frame size, or the use of segment-based metrics for evaluation. Exploring the benefits of using larger frame sizes might also lead to improved performance. However, we suspect that there is a relatively low ceiling to predictive performance, which, if close to an AUC of 0.60 as seen in the present study, would suggest a more important role of emotional elicitors than previously anticipated. Considering the extent of previous findings about the relationships between MECs and various extra-musical factors such as personal meaning or the state of being moved, it is possible that only a small proportion of MECs are caused by acoustic and musical elicitors via the psychological mechanisms of brain stem reflex and musical expectation (see Chapter 2). In other words, instead of acoustic, musical, and emotional elicitors equally contributing to the elicitation of MECs through brain stem reflex, musical expectation, and emotional contagion, the low predictive performance of the model could suggest that, in many cases, emotional elicitors are more predictive of the occurrence of MECs. Regardless of which modelling approach is used, future research should seek to empirically validate the models by generating predictions of MECs on a new set of stimuli, and comparing these predictions to reports of MECs collected from new participants.

Another limitation comes from the reliance of the present study on audio feature extraction, notably in terms of perceptual validity. A simplified way to discuss this limitation is by understanding the purposes such features are put to in different disciplines. In music information retrieval, audio features are generally needed to solve computational tasks (such as genre or emotion classification, music similarity quantification, artist identification, etc.) in order to optimise model performance when compared to a relatively objective ground truth. In music psychology, audio features are generally required to understand underlying psychological processes in order to build cognitive models which are evaluated by comparing them to observed behaviour. These distinctions are explored in detail by Aucouturier and Bigand (2012, 2013), and result in different priorities for feature evaluation. In music information retrieval, features are evaluated following a pragmatic process based on whether or not they improve task performance and are computationally simple, whereas in music psychology, feature evaluation is rare, and focuses on whether or not features are interpretable. Some efforts have been made to identify features which approximate the human perception of related psychological constructs, but these are limited by the fact that they are not easily computable for use in other studies (e.g., Aljanaki & Soleymani, 2018; Friberg et al., 2014). It is therefore very common in psychology experiments to extract features using existing music information retrieval toolboxes—as a very crude example, MIRtoolbox (Lartillot et al., 2008) has been cited in more than 600 articles including the word *psychology*—seeing as the need for computational features often exceeds the need for the perceptual validity provided by subjective ratings, as was the case in the present study. Perceptual validation of audio features would greatly improve the interpretability of the present research, and of research in music psychology in general, but such work would represent a considerable undertaking. In the meantime, and despite their advantages, using audio features to model psychophysiological responses such as MECs should come with the understanding that these features should not be equated with the human perception of related constructs.

In summary, we conducted a set of computational analyses, resulting in a large-scale replication of previous findings about acoustic and musical elicitors of MECs, and in the construction of a model which can identify onsets of MECs better than chance. The hypothesised role of expectation in the experience of MECs was confirmed in a series of novel empirical findings, showing that melodic entropy and information content were significantly different between MECs and controls, led to increased predictive performance when included during model training, and were the best predictors of MEC onsets. Future research should seek to explore the many remaining gaps in knowledge about the relationship between expectation and MECs. Notably, questions remain about the exact interaction between uncertainty and surprise, the role of harmonic and rhythmic expectation, the differences between schematic and veridical expectation, and the interactions between elicitors (this chapter), stylistic preference and familiarity (Chapter 3), and affect (Chapter 4).

Chapter 6

Conclusion

6.1 Overview

In this thesis, we presented a body of work aiming to investigate the relationships between MECs, musical expectation, and affective and aesthetic responses, considering research questions about what MECs are, what causes them, and when they occur.

In Chapter 2, we conducted a systematic review of the literature on MECs, allowing us to define them as a fleeting, pleasurable bodily sensation, sometimes accompanied by goosebumps, experienced when listening to specific musical passages. In the review, we integrated theoretical and empirical findings to reveal that MECs are associated with physiological changes and increased arousal, and recruit brain structures and systems relevant to emotion, reward, and motivation. We identified that they can be caused by a set of acoustic, musical, and emotional elicitors, and are influenced by personality differences such as openness to experience. We provided a preliminary theoretical model that allows for different psychological pathways for the experience of MECs, if not different types of MECs, relying on complex interactions between listener, context, elicitors, psychological and evolutionary mechanisms, and response attributes. We provided a dataset of pieces of music known to cause MECs, and a set of open issues, hypotheses, and methodological recommendations, which motivated the research presented in the following chapters. Notably, we highlighted that further evidence was needed on the relationships between MECs and piloerection, pleasure, familiarity, stylistic preference, and musical expectation, and that causal approaches and the use of naturalistic listening experiences should be emphasised in future research.

In Chapter 3, we followed these recommendations by investigating the relationships between MECs, piloerection, pleasure, musical content, stylistic preference, familiarity, and liking in a controlled, longitudinal experiment using existing pieces of music in as naturalistic a listening context as possible. Musical content, stylistic preference, and familiarity were systematically manipulated, allowing us to identify robust effects of stylistic preference on MECs, piloerection, and pleasure, and revealing that much of the variance in liking pieces of music was accounted for by all the other factors combined. There were fundamental differences between MECs and piloerection, with MECs being reported far more often than piloerection was detected, familiarity having opposite effects on these two responses, and the responses themselves being uncorrelated. However, piloerection significantly overlapped with MECs and pleasure, which suggests that, while related, piloerection might only be present for a distinct type of MECs, or that it requires MECs to exceed an intensity threshold. Effect sizes were small for all the identified effects, and since no effect of musical content on MECs was detected, we hypothesised that more powerful approaches were needed to identify such an effect.

In Chapter 4, we implemented such an approach by conducting a corpus analysis using computational methods. We compared track-level audio features between tracks taken from the dataset provided in Chapter 2 and several sets of control tracks, algorithmically matched by artist, duration, and popularity, to investigate the effect of expressed valence on MECs. We identified that tracks known to elicit MECs were sadder than control tracks, and that they also tended to be slower, less intense, and more instrumental. Moreover, that effect of valence on MECs differed depending on the audio characteristics of the tracks taken from the dataset, suggesting the possibility of there being different causes of MECs for different types of music. Overall, this study provided further evidence for an effect of valence on MECs, and demonstrated that computational methods are well-suited to the study of MECs, enabling the findings that music that causes MECs differed in musical content from other music. However, track-level features were used, which are inadequate for an in-depth exploration of local elicitors of MECs, which occur transiently at particular points while listening to particular pieces of music. This motivated further, more thorough computational work in the following chapter.

In Chapter 5, we conducted a computational analysis aimed at modelling the onset of MECs based on auditory and musical characteristics. We extracted a wide range of features and labels of onsets of MECs from the results of a survey study initiated in Chapter 3, to investigate the effects of known acoustic and musical elicitors of MECs, as well as the often hypothesised effect of musical expectation on the occurrence of MECs. We ran a series of permutation tests to assess the local behaviour of each feature, and trained a series of models to evaluate how well they could predict onsets of MECs, and which features were most important in driving these predictions. This process resulted in a systematic, large-scale replication of all the effects of the elicitors of MECs that were included in the analysis. In addition, onsets of MECs could be predicted better than chance, and melodic expectation improved model performance and was the best predictor of MECs, with MECs being more likely in uncertain melodic contexts including a unexpected event followed by more expected events. The differences between feature importance and local behaviour for some features allowed us to identify that many acoustic elicitors of MECs might be best considered as necessary but not specifically predictive of experiences of MECs.

6.2 Recurring themes

There were four recurring themes that permeated the work presented in this thesis. First, we highlighted theoretical and practical limitations in prior research on MECs, with issues in terms of research design, adequacy of measures of MECs, reproducibility, generalisability, and ecological validity. We have provided contributions to addressing these issues, by conducting a systematic review of the diverse and fast-growing body of research on MECs, developing a preliminary theoretical model of MECs that provides a robust framework for future hypothesis-driven research, and providing a set of open questions, hypotheses, suggested approaches, and methodological recommendations for future research. We also compiled publicly available datasets to improve reproducibility and provide more representative data, made use of existing pieces in music in causal (and arguably double-blinded in Chapter 3) or highly controlled hypothesis-driven studies, explored the use of computational methods which allowed improvements in generalisability and statistical power, and demonstrated the validity of a stimulus-matching paradigm over several studies.

Second, we hypothesised the presence of different psychological pathways for the experience of MECs. While the evidence we gathered is not sufficient to confirm all the predictions arising from the preliminary model of MECs presented in Chapter 2, since that model represents a much larger research agenda than it was possible to cover empirically in this thesis, we did provide empirical support for several of its components, and were not able to refute any of the predictions made by the model. Notably, we confirmed the effects of many acoustic and musical elicitors of MECs, as well as an effect of valence as an emotional elicitor. We identified that the psychological mechanism of musical expectation was a strong predictor of MECs, which potentially provided more explanatory power than a hypothesised involvement of brain stem reflex. This result is particularly significant, because a relationship between expectation and MECs has been postulated continuously and very prominently in the literature since 1991, but there has been no convincing empirical evidence for a relationship to date. Finally, given the effect sizes we observed in behavioural and computational studies, we suspect that there is a ceiling to the effects of acoustic and musical elicitors on MECs, and that emotional elicitors such as meaning or emotionality might be at the origin of most experiences of MECs, or at least more influential, widespread, and consistent elicitors of MECs than acoustic and musical elicitors, through the process of being moved (for recent evidence of a strong relationship between MECs and being moved, see Vuoskoski et al., 2022).

Third, we discussed the possibility that MECs are a collection of phenomenologically and psychologically distinct experiences, as identified in prior research which is discussed in Chapter 2. Again, while we did not explicitly investigate this research question, and are therefore not able to provide conclusive evidence, our findings certainly provided some degree of support for this hypothesis. In fact, all the studies we conducted resulted in the identification of diverging patterns in the experience of MECs, such as fundamental differences between MECs and piloerection in Chapter 3, differences in the effect of valence on MECs depending on stimulus-driven properties in Chapter 4, or low predictive power of acoustic and musical elicitors of MECs in Chapter 5. The presence of distinct types of MECs has received further support in the recent literature (Bannister, 2020a; Bannister & Eerola, 2021), and we consider this research question crucial to future work on MECs.

Lastly, we mentioned the lack of clarity in the relationship between MECs and emotional and aesthetic responses. In Chapter 2, we characterised MECs as a pleasurable, though not essential component of emotional and aesthetic experiences. In Chapter 3, our findings suggested that MECs overlapped significantly but not exclusively with pleasure, and that MECs, in combination with other factors such as stylistic preference and familiarity, could predict a large amount of the variance in music preference. In Chapter 5, we identified that musical expectation was partially predictive of onsets of MECs—a result which can be placed in the context of prior research revealing that violations of musical expectation can induce emotional and aesthetic responses (see Chapter 1). These results provide further justification for the recommendations we made in Chapter 2 that MECs should not be conflated with peak pleasure, and that while they can form a part of emotional and aesthetic responses to music, they should not be used as the sole indicator of such responses. Interestingly, this view is supported by recent evidence that, following administration of an opioid antagonist, experiences of MECs were characterised by no changes in self-reports of pleasure, but decreased pupil diameter (Laeng et al., 2021), therefore suggesting that the removal of a physiological component of MECs had no effect on experienced pleasure.

6.3 Limitations and future work

We opted to focus on MECs in the present research. However, it is important to keep in mind that pleasurable chills can also occur when presented with other forms of art. While some previous research has investigated such responses (see Chapter 2), notably in comparative evaluations of emotional elicitors of chills, it is currently unknown whether or not the hypothesised psychological pathways for distinct experiences of MECs would apply to other types of art-elicited chills.

Many decisions about study design were taken in order to improve ecological validity, notably through the use of naturalistic stimuli, and to gather as much evidence as possible, by opting for interpretable models at the expense of predictive performance (although interpretability was also affected by the lack of perceptual validation of most of the features extracted for the analyses in Chapters 4 and 5). We believe both of these decisions contributed to the small effect sizes observed throughout the present work, along with the impossibility of producing a comprehensive ground truth and the efforts made to provide rigorous control conditions. We suspect that complementary findings could be obtained from studies recruiting more powerful modelling approaches.

In terms of generalisability, while the datasets used in our research featured music from different genres and cultures, they were still mostly comprised of Western music. Relatedly, the behavioural study conducted in Chapter 3 suffered from the same pitfalls as many other psychology experiments with regards to sample representativeness. Cross-lab work involving online methods, as implemented by Jacoby et al. (2021), is expensive in terms of time and resources, but provides an unparalleled opportunity to bring about generalisable cross-cultural findings, from which research on MECs could certainly benefit.

Finally, while establishing causality was a strong motivation for the present work, we generally focused our modelling efforts on maximising interpretability in order to generate novel findings about elicitors of MECs and demonstrate the suitability of computational approaches in the study of MECs. Empirical verification of the predictions made by these models is a necessary next step in order to gain confidence in such findings, and could be complemented by experiments seeking to manipulate stimuli in order to provide causal evidence for the effects of the identified psychological mechanisms underlying the experience of MECs.

Overall, while MECs are inherently subjective, hard to define and measure, and subject to complex interactions between listener, context, and music, they represent a fascinating opportunity to better understand why and how people appreciate music.

Bibliography

Ajmera, J., McCowan, I., & Bourlard, H. (2003). Speech/music segmentation using entropy and dynamism features in a HMM classification framework. *Speech Communication*, *40*(3), 351–363. https://doi.org/10.1016/S0167-6393(02)00087-0

Aljanaki, A., & Soleymani, M. (2018). A data-driven approach to mid-level perceptual musical feature modeling. In E. Gómez, X. Hu, E. Humphrey, & E. Benetos (Eds.), *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018* (pp. 615–621). IRCAM.

Altenmüller, E., Kopiez, R., & Grewe, O. (2013). A contribution to the evolutionary basis of music: Lessons from the chill response. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of emotional communication: From sounds in nonhuman mammals to speech and music in man* (pp. 313–335). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.003.0019

Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods*, *51*(2), 778–792. https://doi.org/10.3758/s13428-018-1095-7

Archie, P., Bruera, E., & Cohen, L. (2013). Music-based interventions in palliative cancer care: A review of quantitative studies and neurobiological literature. *Supportive Care in Cancer*, *21*, 2609–2624. https://doi.org/10.1007/s00520-013-1841-4

Aucouturier, J.-J., & Bigand, E. (2012). Mel Cepstrum & Ann Ova: The difficult dialog between MIR and music cognition. In F. Gouyon, P. Herrera, L. G. Martins, & M. Müller (Eds.), *Proceedings of the 13th Interna-*

tional Society for Music Information Retrieval Conference, ISMIR 2012 (pp. 397–402). FEUP Edições.

Aucouturier, J.-J., & Bigand, E. (2013). Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *Journal of Intelligent Information Systems*, *41*, 483–497. https://doi.org/10.1007/s10844-013-0251-x

Auricchio, N. (2017). Natural highs: Timbre and chills in electronic dance music. In J. Merrill (Ed.), *Popular music studies today: Proceedings of the International Association for the Study of Popular Music 2017* (pp. 11–23). Springer. https://doi.org/10.1007/978-3-658-17740-9_1

Balteș, R. F., Avram, J., Miclea, M., & Miu, A. C. (2011). Emotions induced by operatic music: Psychophysiological effects of music, plot, and acting: A scientist's tribute to Maria Callas. *Brain and Cognition*, *76*(1), 146–157. https://doi.org/10.1016/j.bandc.2011.01.012

Balteș, R. F., & Miu, A. C. (2014). Emotions during live music performance: Links with individual differences in empathy, visual imagery, and mood. *Psychomusicology: Music, Mind, and Brain*, *24*(1), 58–65. https://doi.org/10.1037/pmu0000030

Bannister, S. (2019). Distinct varieties of aesthetic chills in response to multimedia. *PLOS ONE*, *14*(11), e0224974. https://doi.org/10.1371/journal.pone.0224974

Bannister, S. (2020a). *A framework of distinct musical chills: Theoretical, causal, and conceptual evidence* (Doctoral dissertation). Durham University.

Bannister, S. (2020b). A survey into the experience of musically induced chills: Emotions, situations and music. *Psychology of Music*, *48*(2), 297–314. https://doi.org/10.1177/0305735618798024

Bannister, S. (2020c). A vigilance explanation of musical chills? Effects of loudness and brightness manipulations. *Music & Science*, *3*, 1–17. https://doi.org/10.1177/2059204320915654

Bannister, S., & Eerola, T. (2017). *Suppressing the chills: Self-reports, physiological and psychoacoustic correlates (V2)*. Harvard Dataverse. https://doi.org/10.7910/dvn/iucn1q

Bannister, S., & Eerola, T. (2018). Suppressing the chills: Effects of musical manipulation on the chills response. *Frontiers in Psychology*, *9*, 2046. https://doi.org/10.3389/fpsyg.2018.02046

Bannister, S., & Eerola, T. (2021). Vigilance and social chills with music: Evidence for two types of musical chills. *Psychology of Aesthetics, Creativity, and the Arts*, *Advance online publication*. https://doi.org/10.1037/aca0000421

Bartoń, K. (2020). *MuMIn: Multi-model inference (R package version 1.43.17)*. https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beier, E. J., Janata, P., Hulbert, J. C., & Ferreira, F. (2020). Do you chill when I chill? A cross-cultural study of strong emotional responses to music. *Psychology of Aesthetics, Creativity, and the Arts*, *Advance online publication*. https://doi.org/10.1037/aca0000310

Benedek, M., & Kaernbach, C. (2011). Physiological correlates and emotional specificity of human piloerection. *Biological Psychology*, *86*(3), 320–329. https://doi.org/10.1016/j.biopsycho.2010.12.012

Benedek, M., Wilfling, B., Lukas-Wolfbauer, R., Katzur, B. H., & Kaernbach, C. (2010). Objective and continuous measurement of piloerection. *Psychophysiology*, *47*(5), 989–993. https://doi.org/10.1111/j.1469-8986.2010.01003.x

Bériachvili, G. (2016). Frisson esthétique: À la recherche d'une explication théorique. *International Review of the Aesthetics and Sociology of Music*, *47*(1), 63–85.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.

Bharucha, J. J. (1994). Tonality and expectation. In R. Aiello & J. A. Sloboda (Eds.), *Musical perceptions* (pp. 213–239). Oxford University Press.

Bicknell, J. (2007). Explaining strong emotional responses to music: Sociality and intimacy. *Journal of Consciousness Studies*, *14*(12), 5–23.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Bonneville-Roussy, A., Stillwell, D., Kosinski, M., & Rust, J. (2017). Age trends in musical preferences in adulthood: I. Conceptualization and empirical investigation. *Musicae Scientiae*, *21*(4), 369–389. https://doi.org/10.1177/1029864917691571

Branković, S. (2013). Neuroaesthetics and growing interest in "positive affect" in psychiatry: New evidence and prospects for the theory of informational needs. *Psychiatria Danubina*, *25*(2), 97–107.

Brattico, E. (2019). The neuroaesthetics of music: A research agenda coming of age. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 364–390). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198804123.013.15

Brattico, E., Bogert, B., & Jacobsen, T. (2013). Toward a neural chronometry for the aesthetic experience of music. *Frontiers in Psychology*, *4*, 206. https://doi.org/10.3389/fpsyg.2013.00206

Brattico, E., Brattico, P., & Jacobsen, T. (2009). The origins of the aesthetic enjoyment of music — A review of the literature. *Musicae Scientiae*, *13*(Suppl. 2), 15–39. https://doi.org/10.1177/1029864909013002031

Brattico, E., & Pearce, M. T. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48–61. https://doi.org/10.1037/a0031624

Brattico, E., & Varankaitė, U. (2019). Aesthetic empowerment through music. *Musicae Scientiae*, *23*(3), 285–303. https://doi.org/10.1177/1029864919850606

Cantor, R. M. (2019). On the embodied meaning of emotional responses to music: A semiotic perspective. *Semiotica*, *231*, 225–244. https://doi.org/10.1515/sem-2018-0039

Carr, S. M., & Rickard, N. S. (2016). The use of emotionally arousing music to enhance memory for subsequently presented images. *Psychology of Music*, *44*(5), 1145–1157. https://doi.org/10.1177/0305735615613846

Chabin, T., Tio, G., Comte, A., Joucla, C., Gabriel, D., & Pazart, L. (2020). The relevance of a conductor competition for the study of emotional synchronization within and between groups in a natural music setting. *Frontiers in Psychology*, *10*, 2954. https://doi.org/10.3389/fpsyg.2019.02954

Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. *Trends in Cognitive Sciences*, *17*(4), 179–193. https://doi.org/10.1016/j.tics.2013.02.007

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. *Current Biology*, *29*(23), 4084–4092. https://doi.org/10.1016/j.cub.2019.09.067

Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature. *Psychology of Music*, *45*(6), 886–909. https://doi.org/10.1177/0305735617697507

Chmiel, A., & Schubert, E. (2019). Unusualness as a predictor of music preference. *Musicae Scientiae*, *23*(4), 426–441. https://doi.org/10.1177/1029864917752545

Colver, M. C., & El-Alayli, A. (2016). Getting aesthetic chills from music: The connection between openness to experience and frisson. *Psychology of Music*, *44*(3), 413–427. https://doi.org/10.1177/0305735615572358

Corrigall, K. A., & Schellenberg, E. G. (2013). Music: The language of emotion. In C. Mohiyeddini, M. Eysenck, & S. Bauer (Eds.), *Handbook of psychology of emotions: Recent theoretical perspectives and novel empirical findings* (pp. 299–326). Nova Science Publishers.

Corrigall, K. A., & Schellenberg, E. G. (2015). Liking music: Genres, contextual factors, and individual differences. In J. P. Huston, M. Nadal, F. Mora, L. F. Agnati, & C. J. C. Conde (Eds.), *Art, aesthetics, and the brain* (pp. 263–284). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199670000.003.0013

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.

Cotter, K. N., Prince, A. N., Christensen, A. P., & Silvia, P. J. (2019). Feeling like crying when listening to music: Exploring musical and contextual features. *Empirical Studies of the Arts*, *37*(2), 119–137. https://doi.org/10.1177/0276237418805692

Cotter, K. N., Silvia, P. J., & Fayn, K. (2018). What does feeling like crying when listening to music feel like? *Psychology of Aesthetics, Creativity, and the Arts*, *12*(2), 216–227. https://doi.org/10.1037/aca0000108

Craig, D. G. (2005). An exploratory study of physiological changes during "chills" induced by music. *Musicae Scientiae*, *9*(2), 273–287. https://doi.org/10.1177/102986490500900207

Craig, D. G. (2009). Exploring music preference: Meaningfulness of music as a function of emotional reactions. *Nordic Journal of Music Therapy*, *18*(1), 57–69. https://doi.org/10.1080/08098130802697137

de Fleurian, R., & Pearce, M. T. (2021). The relationship between valence and chills in music: A corpus analysis. *i-Perception*, *12*(4), 1–11. https://doi.org/10.1177/20416695211024680

del Campo, M. A., & Kehle, T. J. (2016). Autonomous sensory meridian response (ASMR) and frisson: Mindfully induced sensory phenomena that promote happiness. *International Journal of School & Educational Psychology*, *4*(2), 99–105. https://doi.org/10.1080/21683603.2016.1130582

Dowd, E. C., & Barch, D. M. (2012). Pavlovian reward prediction and receipt in schizophrenia: Relationship to anhedonia. *PLOS ONE*, *7*(5), e35622. https://doi.org/10.1371/journal.pone.0035622

Dubé, L., & Le Bel, J. (2003). The content and structure of laypeople's concept of pleasure. *Cognition and Emotion*, *17*(2), 263–295. https://doi.org/10.1080/02699930302295

Dushoff, J., Kain, M. P., & Bolker, B. M. (2019). I can see clearly now: Reinterpreting statistical significance. *Methods in Ecology and Evolution*, *10*(6), 756–759. https://doi.org/10.1111/2041-210X.13159

Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, *40*(4), 349–366. https://doi.org/10.1080/09298215.2011.602195

Eerola, T. (2018). Music and emotions. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 539–554). Springer. https://doi.org/10.1007/978-3-662-55004-5_29

Eerola, T., Vuoskoski, J. K., & Kautiainen, H. (2016). Being moved by unfamiliar sad music is associated with high empathy. *Frontiers in Psychology*, *7*, 1176. https://doi.org/10.3389/fpsyg.2016.01176

Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, and Behavioral Neuroscience*, *13*(3), 533–553. https://doi.org/10.3758/s13415-013-0161-y

Egermann, H., Sutherland, M. E., Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2011). Does music listening in a social context alter experience? A physiological and psychological perspective on emotion. *Musicae Scientiae*, *15*(3), 307–323. https://doi.org/10.1177/1029864911399497

Fechner, G. T. (1876). *Vorschule der Ästhetik*. Breitkopf & Härtel.

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G., Marco-Pallarés, J., Antonijoan, R. M., Valle, M., Riba, J., & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(9), 3793–3798. https://doi.org/10.1073/pnas.1811878116

Fiske, A. P., Seibt, B., & Schubert, T. W. (2019). The sudden devotion emotion: Kama muta and the cultural practices whose function is to evoke it. *Emotion Review*, *11*(1), 74–86. https://doi.org/10.1177/1754073917723167

Friberg, A., Schoonderwaldt, E., Hedblad, A., Fabiani, M., & Elowsson, A. (2014). Using listener-based perceptual features as intermediate representations in music information retrieval. *The Journal of the Acoustical Society of America*, *136*, 1951–1963. https://doi.org/10.1121/1.4892767

Fukui, H., & Toyoshima, K. (2013). Influence of music on steroid hormones and the relationship between receptor polymorphisms and musical ability: A pilot study. *Frontiers in Psychology*, *4*, 910. https://doi.org/10.3389/fpsyg.2013.00910

Fukui, H., & Toyoshima, K. (2014). Chill-inducing music enhances altruism in humans. *Frontiers in Psychology*, *5*, 1215. https://doi.org/10.3389/fpsyg.2014.01215

Fukushima, S., & Kajimoto, H. (2012). Facilitating a surprised feeling by artificial control of piloerection on the forearm. In J.-M. Seigneur (Ed.), *Proceedings of the 3rd Augmented Human International Conference, AH '12* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/2160125.2160133

Gabrielsson, A. (2001a). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, *5*(Suppl. 1), 123–147. https://doi.org/10.1177/10298649020050s105

Gabrielsson, A. (2001b). Emotions in strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 431–449). Oxford University Press.

Gabrielsson, A. (2011). *Strong experiences with music: Music is much more than just music*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199695225.001.0001

Gabrielsson, A., Whaley, J., & Sloboda, J. (2016). Peak experiences in music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd, pp. 745–758). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.44

Gabrielsson, A., & Wik, S. L. (2003). Strong experiences related to music: A descriptive system. *Musicae Scientiae*, *7*(2), 157–217. https://doi.org/10.1177/102986490300700201

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, *37*(2), 93–100. https://doi.org/10.1080/09298210802479268

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *Journal of Neuroscience*, *39*(47), 9397–9409. https://doi.org/10.1523/jneurosci.0428-19.2019

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology*, *8*(1), 126–129. https://doi.org/10.3758/bf03326460

Gómez, E. (2006). Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, *18*(3), 294–304. https://doi.org/10.1287/ijoc.1040.0126

Goodchild, M., Wild, J., & McAdams, S. (2019). Exploring emotional responses to orchestral gestures. *Musicae Scientiae*, *23*(1), 25–49. https://doi.org/10.1177/1029864917704033

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Greasley, A., & Lamont, A. (2016). Musical preferences. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd, pp. 263–281). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.58

Grewe, O., Katzur, B., Kopiez, R., & Altenmüller, E. (2011). Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*, *39*(2), 220–239. https://doi.org/10.1177/0305735610362950

Grewe, O., Kopiez, R., & Altenmüller, E. (2009a). The chill parameter: Goose bumps and shivers as promising measures in emotion research. *Music Perception*, *27*(1), 61–74. https://doi.org/10.1525/mp.2009.27.1.61

Grewe, O., Kopiez, R., & Altenmüller, E. (2009b). Chills as an indicator of individual emotional peaks. *Annals of the New York Academy of Sciences*, *1169*(1), 351–354. https://doi.org/10.1111/j.1749-6632.2009.04783.x

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2007). Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, *24*(3), 297–314. https://doi.org/10.1525/mp.2007.24.3.297

Griffiths, T. D., Warren, J. D., Dean, J. L., & Howard, D. (2004). "When the feeling's gone": A selective loss of musical emotion. *Journal of Neurology, Neurosurgery, and Psychiatry*, *72*(2), 344–345. https://doi.org/10.1136/jnnp.2003.015586

Grunkina, V., Holtz, K., Klepzig, K., Neubert, J., Horn, U., Domin, M., Hamm, A. O., & Lotze, M. (2017). The role of left hemispheric structures for emotional processing as a monitor of bodily reaction and felt chill – A case-control functional imaging study. *Frontiers in Human Neuroscience*, *10*, 670. https://doi.org/10.3389/fnhum.2016.00670

Guhn, M., Hamm, A., & Zentner, M. (2007). Physiological and musico-acoustic correlates of the chill response. *Music Perception*, *24*(5), 473–483. https://doi.org/10.1525/mp.2007.24.5.473

Gulmatico, J. S., Susa, J. A. B., Malbog, M. A. F., Acoba, A., Nipas, M. D., & Mindoro, J. N. (2022). SpotiPred: A machine learning approach prediction of Spotify music popularity by audio features. In P. D. Dewangan, S. Ghosh, R. N. Patel, V. H. Priya, & P. Chaturvedi (Eds.), *Proceedings of the 2022 Second International Conference on Power, Control and Computing Technologies, ICPC2T* (pp. 1–5). IEEE. https://doi.org/10.1109/icpc2t53885.2022.9776765

Habibi, A., & Damasio, A. (2014). Music, feelings, and the human brain. *Psychomusicology: Music, Mind, and Brain*, *24*(1), 92–102. https://doi.org/10.1037/pmu0000033

Halpern, D. L., Blake, R., & Hillenbrand, J. (1986). Psychoacoustics of a chilling sound. *Perception & Psychophysics*, *39*(2), 77–80. https://doi.org/10.3758/bf03211488

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*, 1052. https://doi.org/10.3389/fpsyg.2014.01052

Hansen, N. C., Vuust, P., & Pearce, M. T. (2016). "If you have to ask, you'll never know": Effects of specialised stylistic expertise on predictive processing of music. *PLOS ONE*, *11*(10), e0163584. https://doi.org/10.1371/journal.pone.0163584

Hanslick, E. (1854). *Vom Musikalisch-Schönen: Ein Beitrag zur Revision der Ästhetik der Tonkunst*. Rudolph Weigel.

Hargreaves, D. J. (2012). Musical imagination: Perception and production, beauty and creativity. *Psychology of Music*, *40*(5), 539–557. https://doi.org/10.1177/0305735612444893

Hargreaves, D. J., & North, A. C. (2010). Experimental aesthetics and liking for music. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 515–546). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0019

Harrison, L., & Loui, P. (2014). Thrills, chills, frissons, and skin orgasms: Toward an integrative model of transcendent psychophysiological experiences in music. *Frontiers in Psychology*, *5*, 790. https://doi.org/10.3389/fpsyg.2014.00790

Harrison, N. R., & Clark, D. P. A. (2016). The observing facet of trait mindfulness predicts frequency of aesthetic experiences evoked by the arts. *Mindfulness*, *7*, 971–978. https://doi.org/10.1007/s12671-016-0536-6

Harrison, P. M. C. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*, *5*(49), 2088. https://doi.org/10.21105/joss.02088

Harvey, P.-O., Pruessner, J., Czechowska, Y., & Lepage, M. (2007). Individual differences in trait anhedonia: A structural and functional magnetic resonance imaging study in non-clinical subjects. *Molecular Psychiatry*, *12*, 767–775. https://doi.org/10.1038/sj.mp.4002021

Heathers, J. A. J., Fayn, K., Silvia, P. J., Tiliopoulos, N., & Goodwin, M. S. (2018). The voluntary control of piloerection. *PeerJ*, *6*, e5292. https://doi.org/10.7717/peerj.5292

Helmholtz, H. v. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. J. Vieweg.

Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, *5*(50), 2154. https://doi.org/10.21105/joss.02154

Hernández, M., Palomar-García, M.-Á., Nohales-Nieto, B., Olcina-Sempere, G., Villar-Rodríguez, E., Pastor, R., Ávila, C., & Parcet, M.-A. (2019). Separate contribution of striatum volume and pitch discrimination to individual differences in music reward. *Psychological Science*, *30*(9), 1352–1361. https://doi.org/10.1177/0956797619859339

Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, *17*(1), 84–90. https://doi.org/10.3758/bf03204003

Hodges, D. A. (2016). Bodily responses to music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd, pp. 183–196). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.16

Honda, S., Ishikawa, Y., Konno, R., Imai, E., Nomiyama, N., Sakurada, K., Koumura, T., Kondo, H. M., Furukawa, S., Fujii, S., & Nakatani, M. (2020). Proximal binaural sound can induce subjective frisson. *Frontiers in Psychology*, *11*, 316. https://doi.org/10.3389/fpsyg.2020.00316

Hunter, P. G., & Schellenberg, E. G. (2010). Music and emotion. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (pp. 129–164). Springer. https://doi.org/10.1007/978-1-4419-6114-3_5

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Huron, D. (2016). Aesthetics. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd, pp. 233–245). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.19

Huron, D., & Margulis, E. H. (2010). Musical expectancy and thrills. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 575–604). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0021

Ishikawa, Y., Kawazoe, A., Chernyshov, G., Fujii, S., & Nakatani, M. (2019). The thermal feedback influencer: Wearable thermal display for enhancing the experience of music listening. In H. Kajimoto, D. Lee, S.-Y. Kim, M. Konyo, & K.-U. Kyung (Eds.), *Proceedings of the International*

AsiaHaptics Conference, AsiaHaptics 2018: Haptic interaction (pp. 162–168). Springer. https://doi.org/10.1007/978-981-13-3194-7_36

Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Goldberg, D., Margulis, E., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., … McDermott, J. H. (2021). Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors. *PsyArXiV*. https://doi.org/10.31234/osf.io/b879v

Jaimovich, J., Coghlan, N., & Knapp, R. B. (2013). Emotion in motion: A study of music and affective response. In M. Aramaki, M. Barthet, R. Kronland-Martinet, & S. Ystad (Eds.), *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval, CMMR 2012: From sounds to music and emotions* (pp. 19–43). Springer. https://doi.org/10.1007/978-3-642-41248-6_2

Jakubowski, K., Finkel, S., Stewart, L., & Müllensiefen, D. (2017). Dissecting an earworm: Melodic features and song popularity predict involuntary musical imagery. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(2), 122–135. https://doi.org/10.1037/aca0000090

Jensenius, A. R. (2002). *How do we recognize a song in one second?* (Doctoral dissertation). University of Oslo.

Ji, Q., Janicke-Bowles, S. H., De Leeuw, R. N. H., & Oliver, M. B. (2019). The melody to inspiration: The effects of awe-eliciting music on approach motivation and positive well-being. *Media Psychology*, *24*(3), 305–331. https://doi.org/10.1080/15213269.2019.1693402

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491. https://doi.org/10.1037/0033-295x.96.3.459

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics*, *33*(3), 251–272. https://doi.org/10.1080/00401706.1991.10484833

Jurafsky, D., & Martin, J. H. (2021). Hidden Markov models. *Speech and language processing* (pp. 1–17).

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235–266. https://doi.org/10.1016/j.plrev.2013.05.008

Juslin, P. N. (2016). Emotional reactions to music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd, pp. 197–213). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.17

Juslin, P. N. (2019). *Musical emotions explained*. Oxford University Press. https://doi.org/10.1093/oso/9780198753421.001.0001

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*(4), 599–623. https://doi.org/10.1177/0305735613484548

Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L.-O. (2010). How does music evoke emotions? In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 605–642). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0022

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(5), 559–575. https://doi.org/10.1017/s0140525x08005293

Kim, J., Seo, D. G., & Cho, Y.-H. (2014). A flexible skin piloerection monitoring sensor. *Applied Physics Letters*, *104*, 253502. https://doi.org/10.1063/1.4881888

Klapuri, A. P., Eronen, A. J., & Astola, J. T. (2006). Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(1), 342–355. https://doi.org/10.1109/TSA.2005.854090

Klepzig, K., Horn, U., König, J., Holtz, K., Wendt, J., Hamm, A. O., & Lotze, M. (2020). Brain imaging of chill reactions to pleasant and unpleasant sounds. *Behavioural Brain Research*, *380*, 112417. https://doi.org/10.1016/j.bbr.2019.112417

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, *14*(3), 131–137. https://doi.org/10.1016/j.tics.2010.01.002

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*, 170–180. https://doi.org/10.1038/nrn3666

Koelsch, S., & Jäncke, L. (2015). Music and the heart. *European Heart Journal*, *36*, 3043–3048. https://doi.org/10.1093/eurheartj/ehv430

Konečni, V. J. (1979). Determinants of aesthetic preference and effects of exposure to aesthetic stimuli: Social, emotional, and cognitive factors. In B. A. Maher (Ed.), *Progress in experimental personality research* (pp. 149–197). Academic Press.

Konečni, V. J. (2005). The aesthetic trinity: Awe, being moved, thrills. *Bulletin of Psychology and the Arts*, *5*(2), 27–44. https://doi.org/10.1037/e674862010-005

Konečni, V. J. (2007). Music and emotion: An empirical critique of a key issue in the philosophy of music. In A. Š. et al. (Ed.), *Proceedings of the 5th International Conference "Person – Color – Nature – Music"* (pp. 31–40). Daugavpils University.

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 115–129. https://doi.org/10.1037/1931-3896.2.2.115

Konečni, V. J. (2013). Empirical psycho-aesthetics and her sisters: Substantive and methodological issues—Part II. *The Journal of Aesthetic Education*, *47*(1), 1–21. https://doi.org/10.5406/jaesteduc.47.1.0001

Konečni, V. J., Wanic, R. A., & Brown, A. (2007). Emotional and aesthetic antecedents and consequences of music-induced thrills. *American Journal of Psychology*, *120*(4), 619–643. https://doi.org/10.2307/20445428

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception*, *7*(3), 309–324. https://doi.org/10.2307/40285467

Krumhansl, C. L. (2010). Plink: "thin slices" of music. *Music Perception*, *27*(5), 337–354. https://doi.org/10.1525/mp.2010.27.5.337

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laeng, B., Eidet, L. M., Sulutvedt, U., & Panksepp, J. (2016). Music chills: The eye pupil as a mirror to music's soul. *Consciousness and Cognition*, *44*, 161–178. https://doi.org/10.1016/j.concog.2016.07.009

Laeng, B., Garvija, L., Løseth, G., Eikemo, M., Ernst, G., & Leknes, S. (2021). 'Defrosting' music chills with naltrexone: The role of endogenous opioids for the intensity of musical pleasure. *Consciousness and Cognition*, *90*, 103105. https://doi.org/10.1016/j.concog.2021.103105

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. https://doi.org/10.1037/0033-295x.106.1.119

Larsen, J. T., Berntson, G. G., Poehlmann, K. M., Ito, T. A., & Cacioppo, J. T. (2008). The psychophysiology of emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd, pp. 180–195). Guilford Press.

Lartillot, O. (2019). MiningSuite: A comprehensive Matlab framework for signal, audio, and music analysis, articulating audio and symbolic approaches. In I. Barbancho, L. J. Tardón, A. Peinado, & A. M. Barbancho (Eds.), *SMC 2019 Proceedings of the 16th Sound & Music Computing Conference* (p. 489). Universidad de Málaga.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications: Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation* (pp. 261–268). Springer. https://doi.org/10.1007/978-3-540-78246-9_31

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *94*(4), 489–508. https://doi.org/10.1348/0007126042369811

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode – Developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*(4), 443–464. https://doi.org/10.1111/bjop.12084

Lee, E.-J. (2008). The thrill effect in medical treatment: Thrill effect as a therapeutic tool in clinical health care (esp. music therapy). In S.-D. Yoo (Ed.), *Proceedings of the EU-Korea Conference on Science and Technology. EKC2008* (pp. 477–483). Springer.

Levinson, J. (2006). Musical chills. In J. Levinson (Ed.), *Contemplating art: Essays in aesthetics* (pp. 220–236). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199206179.003.0013

Levinson, J. (2009). The aesthetic appreciation of music. *British Journal of Aesthetics*, *49*(4), 415–425. https://doi.org/10.1093/aesthj/ayp043

Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2018). *tuneR: Analysis of music and speech (R package version 1.3.3.)* https://CRAN.R-project.org/package=tuneR

Linnemann, A., Kreutz, G., Gollwitzer, M., & Nater, U. M. (2018). Validation of the German version of the Music-Empathizing-Music-Systemizing (MEMS) inventory (short version). *Frontiers in Behavioral Neuroscience*, *12*, 153. https://doi.org/10.3389/fnbeh.2018.00153

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*, 869–878. https://doi.org/10.1038/nature06976

Loui, P., Patterson, S., Sachs, M. E., Leung, Y., Zeng, T., & Przysinda, E. (2017). White matter correlates of musical anhedonia: Implications for evolution of music. *Frontiers in Psychology*, *8*, 1664. https://doi.org/10.3389/fpsyg.2017.01664

Mace, S. T., Wagoner, C. L., Teachout, D. J., & Hodges, D. A. (2012). Genre identification of very brief musical excerpts. *Psychology of Music*, *40*(1), 112–128. https://doi.org/10.1177/0305735610391347

Madison, G., & Schiölde, G. (2017). Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music. *Frontiers in Neuroscience*, *11*, 147. https://doi.org/10.3389/fnins.2017.00147

Madsen, C. K., Brittin, R. V., & Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, *41*(1), 57–69. https://doi.org/10.2307/3345480

Mallik, A., Chanda, M. L., & Levitin, D. J. (2017). Anhedonia to music and mu-opioids: Evidence from the administration of naltrexone. *Scientific Reports*, *7*, 41952. https://doi.org/10.1038/srep41952

Maruskin, L. A., Thrash, T. M., & Elliot, A. J. (2012). The chills as a psychological construct: Content universe, factor structure, affective composition, elicitors, trait antecedents, and consequences. *Journal of Personality and Social Psychology*, *103*(1), 135–157. https://doi.org/10.1037/a0028117

Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2018). Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. *Nature Human Behaviour*, *2*, 27–32. https://doi.org/10.1038/s41562-017-0241-z

Mas-Herrero, E., Marco-Pallarés, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual differences in music reward experiences. *Music Perception*, *31*(2), 118–138. https://doi.org/10.1525/mp.2013.31.2.118

Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., & Marco-Pallarés, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology*, *24*(6), 699–704. https://doi.org/10.1016/j.cub.2014.01.068

Mathis, W. S., & Han, X. (2017). The acute effect of pleasurable music on craving for alcohol: A pilot crossover study. *Journal of Psychiatric Research*, *90*, 143–147. https://doi.org/10.1016/j.jpsychires.2017.04.008

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7

McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience. *Motivation and Emotion*, *31*(1), 5–11. https://doi.org/10.1007/s11031-007-9053-1

McDermott, J. H. (2012). Auditory preferences and aesthetics: Music, voices, and everyday sounds. In R. J. Dolan & T. Sharot (Eds.), *Neuroscience of* *preference and choice: Cognitive and neural mechanisms* (pp. 227–256). Academic Press. https://doi.org/10.1016/b978-0-12-381431-9.00020-6

McEvilly, D. K. (1999). Chills and tempo. *Music Perception*, *16*(4), 457–462. https://doi.org/10.2307/40285804

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *336*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

Melchiorre, A. B., & Schedl, M. (2020). Personality correlates of music audio preferences for modelling music listeners. In T. Kuflik & I. Torre (Eds.), *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '20* (pp. 313–317). Association for Computing Machinery. https://doi.org/10.1145/3340631.3394874

Mencke, I., Omigie, D., Wald-Fuhrmann, M., & Brattico, E. (2019). Atonal music: Can uncertainty lead to pleasure? *Frontiers in Neuroscience*, *12*, 979. https://doi.org/10.3389/fnins.2018.00979

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PLOS ONE*, *10*(6), e0128451. https://doi.org/10.1371/journal.pone.0128451

Mesaros, A., Heittola, T., & Virtanen, T. (2016). Metrics for polyphonic sound event detection. *Applied Sciences*, *6*(6), 162. https://doi.org/10.3390/app6060162

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago University Press. https://doi.org/10.7208/chicago/9780226521374.001.0001

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, *15*(4), 412–424. https://doi.org/10.2307/427154

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage*, *38*(2), 331–345. https://doi.org/10.1016/j.neuroimage.2007.07.034

Mlejnek, R. (2013). Physically experienced reactions and music: A questionnaire study of musicians and non-musicians. In G. Luck & O. Brabant (Eds.), *Proceedings of the 3rd International Conference on Music and Emotion, ICME3*. University of Jyväskylä, Department of Music.

Mori, K. (2022). Decoding peak emotional responses to music from computational and lyrical features. *Cognition*, *222*, 105010. https://doi.org/10.1016/j.cognition.2021.105010

Mori, K., & Iwanaga, M. (2014a). Music-induced chills as a strong emotional experience. *The Japanese Journal of Psychology*, *85*(5), 495–509. https://doi.org/10.4992/jjpsy.85.13401

Mori, K., & Iwanaga, M. (2014b). Resting physiological arousal is associated with the experience of music-induced chills. *International Journal of Psychophysiology*, *93*(2), 220–226. https://doi.org/10.1016/j.ijpsycho.2014.05.001

Mori, K., & Iwanaga, M. (2015). General reward sensitivity predicts intensity of music-evoked chills. *Music Perception*, *32*(5), 484–492. https://doi.org/10.1525/mp.2015.32.5.484

Mori, K., & Iwanaga, M. (2017). Two types of peak emotional responses to music: The psychophysiology of chills and tears. *Scientific Reports*, *7*, 46063. https://doi.org/10.1038/srep46063

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, *39*(2), 283–290. https://doi.org/10.3758/bf03193159

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2008). Psychoacoustical correlates of musically induced chills. *Musicae Scientiae*, *12*(1), 101–113. https://doi.org/10.1177/102986490801200106

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Neidlinger, K., Truong, K. P., Telfair, C., Feijs, L., Dertien, E., & Evers, V. (2017). AWElectric: That gave me goosebumps, did you feel it too? In R. Peiris (Ed.), *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction, TEI '17* (pp. 315–324). Association for Computing Machinery. https://doi.org/10.1145/3024969.3025004

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology*, *14*(1–2), 77–93. https://doi.org/10.1037/h0094090

Nusbaum, E. C., & Silvia, P. J. (2011). Shivers and timbres: Personality and the experience of chills from music. *Social Psychological and Personality Science*, *2*(2), 199–204. https://doi.org/10.1177/1948550610386810

Nusbaum, E. C., Silvia, P. J., Beaty, R. E., Burgin, C. J., Hodges, D. A., & Kwapil, T. R. (2014). Listening between the notes: Aesthetic chills in everyday music listening. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(1), 104–109. https://doi.org/10.1037/a0034867

Nusbaum, E. C., Silvia, P. J., Beaty, R. E., Burgin, C. J., & Kwapil, T. R. (2015). Turn that racket down! Physical anhedonia and diminished pleasure from music. *Empirical Studies of the Arts*, *33*(2), 228–243. https://doi.org/10.1177/0276237415597392

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783. https://doi.org/10.1037/h0043424

Orr, M. G., & Ohlsson, S. (2005). Relationship between complexity and liking as a function of expertise. *Music Perception*, *22*(4), 583–611. https://doi.org/10.1525/mp.2005.22.4.583

Panksepp, J. (1995). The emotional sources of "chills" induced by music. *Music Perception*, *13*(2), 171–207. https://doi.org/10.2307/40285693

Panksepp, J. (2009). The emotional antecedents to the evolution of music and language. *Musicae Scientiae*, *13*(Suppl. 2), 229–259. https://doi.org/10.1177/1029864909013002111

Panksepp, J., & Bernatzky, G. (2002). Emotional sounds and the brain: The neuro-affective foundations of musical appreciation. *Behavioural Processes*, *60*(2), 133–155. https://doi.org/10.1016/s0376-6357(02)00080-3

Panzarella, R. (1980). The phenomenology of aesthetic peak experiences. *Journal of Humanistic Psychology*, *20*(1), 69–85. https://doi.org/10.1177/002216788002000105

Park, K. S., Hass, C. J., Fawver, B., Lee, H., & Janelle, C. M. (2019). Emotional states influence forward gait during music listening based on familiarity with music selections. *Human Movement Science*, *66*, 53–62. https://doi.org/10.1016/j.humov.2019.03.004

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (Doctoral dissertation). City University of London.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, *4*(4), 625–652. https://doi.org/10.1111/j.1756-8765.2012.01214.x

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me... delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, *21*, 80–125. https://doi.org/10.1016/j.plrev.2017.02.003

Pelowski, M., Markey, P. S., & Leder, H. (2018). Chills, aesthetic experience, and new versus old knowledge — What do chills actually portend?

Comment on "Physics of mind: Experimental confirmations of theoretical predictions" by Schoeller et al. *Physics of Life Reviews*, *25*, 83–87. https://doi.org/10.1016/j.plrev.2018.03.014

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: Familiarity matters. *PLOS ONE*, *6*(11), e27241. https://doi.org/10.1371/journal.pone.0027241

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large-margin classifiers* (pp. 61–74). MIT Press.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Polo, M. J. (2017). *Chill responses to post-tonal music: Musical structures and physiological reactions* (Doctoral dissertation). University of Florida.

Quesnel, D., & Riecke, B. E. (2018). Are you awed yet? How virtual reality gives us awe and goose bumps. *Frontiers in Psychology*, *9*, 2158. https://doi.org/10.3389/fpsyg.2018.02158

Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., & Vuust, P. (2019). Musical prediction error responses similarly reduced by predictive uncertainty in musicians and non-musicians. *European Journal of Neuroscience*, *51*(11), 2250–2269. https://doi.org/10.1111/ejn.14667

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, *100*(6), 1139–1157. https://doi.org/10.1037/a0022406

Rentfrow, P. J., Goldberg, L. R., Stillwell, D. J., Kosinski, M., Gosling, S. D., & Levitin, D. J. (2012). The song remains the same: A replication and extension of the MUSIC model. *Music Perception*, *30*(2), 161–185. https://doi.org/10.1525/mp.2012.30.2.161

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256. https://doi.org/10.1037/0022-3514.84.6.1236

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, *32*(4), 371–388. https://doi.org/10.1177/0305735604046096

Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, *83*(2), 164–175. https://doi.org/10.1016/j.ijpsycho.2011.12.010

Sachs, M. E., Ellis, R. J., Schlaug, G., & Loui, P. (2016). Brain connectivity reflects human aesthetic responses to music. *Social Cognitive and Affective Neuroscience*, *11*(6), 884–891. https://doi.org/10.1093/scan/nsw009

Sachs, M. E., Habibi, A., & Damasio, H. (2018). Reflections on music, affect, and sociality. In J. F. Christensen & A. Gomila (Eds.), *The arts and the brain: Psychology and physiology beyond pleasure* (pp. 153–172). Academic Press. https://doi.org/10.1016/bs.pbr.2018.03.009

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/s0010-0277(98)00075-4

Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(6), 1759–1770. https://doi.org/10.1109/TASL.2012.2188515

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLOS ONE*, *4*(10), e7487. https://doi.org/10.1371/journal.pone.0007487

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. *Trends in Cognitive Sciences*, *19*(2), 86–91. https://doi.org/10.1016/j.tics.2014.12.001

Salimpoor, V. N., & Zatorre, R. J. (2013). Neural interactions that give rise to musical pleasure. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 62–75. https://doi.org/10.1037/a0031819

Satoh, M., Kato, N., Tabei, K.-i., Nakano, C., Abe, M., Fujita, R., Kida, H., Tomimoto, H., & Kondo, K. (2016). A case of musical anhedonia due to right putaminal hemorrhage: A disconnection syndrome between the auditory cortex and insula. *Neurocase*, *22*(6), 518–525. https://doi.org/10.1080/13554794.2016.1264609

Sauvé, S. A., Sayed, A., Dean, R. T., & Pearce, M. T. (2018). Effects of pitch and timing expectancy on musical emotion. *Psychomusicology: Music, Mind, and Brain*, *28*(1), 17–39. https://doi.org/10.1037/pmu0000203

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Schaefer, H.-E. (2017). Music-evoked emotions—Current studies. *Frontiers in Neuroscience*, *11*, 600. https://doi.org/10.3389/fnins.2017.00600

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences*, *116*, 265–273. https://doi.org/10.1016/j.paid.2017.04.061

Schäfer, T., & Sedlmeier, P. (2010). What makes us like music? Determinants of music preference. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(4), 223–234. https://doi.org/10.1037/a0018374

Schäfer, T., & Sedlmeier, P. (2011). Does the body move the soul? The impact of arousal on music preference. *Music Perception*, *29*(1), 37–50. https://doi.org/10.1525/mp.2011.29.1.37

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, *6*(4), 641–646. https://doi.org/10.3758/BF03212973

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*(7), 1307–1351. https://doi.org/10.1080/02699930902928969

Scherer, K. R., Zentner, M. R., & Schacht, A. (2001). Emotional states generated by music: An exploratory study of music experts. *Musicae Scientiae*, *5*(Suppl. 1), 149–171. https://doi.org/10.1177/10298649020050s106

Schoeller, F. (2015). Knowledge, curiosity, and aesthetic chills. *Frontiers in Psychology*, *6*, 1546. https://doi.org/10.3389/fpsyg.2015.01546

Schoeller, F., & Eskinazi, M. (2019). Psychologie du frisson esthétique. *Psychologie Française*, *64*, 305–312. https://doi.org/10.1016/j.psfr.2017.11.003

Schoeller, F., Eskinazi, M., & Garreau, D. (2018). Dynamics of the knowledge instinct: Effects of incoherence on the cognitive system. *Cognitive Systems Research*, *48*, 85–91. https://doi.org/10.1016/j.cogsys.2017.07.005

Schoeller, F., Haar, A. J. H., Jain, A., & Maes, P. (2019). Enhancing human emotions with interoceptive technologies. *Physics of Life Reviews*, *31*, 310–319. https://doi.org/10.1016/j.plrev.2019.10.008

Schoeller, F., & Perlovsky, L. (2016). Aesthetic chills: Knowledge-acquisition, meaning-making, and aesthetic emotions. *Frontiers in Psychology*, *7*, 1093. https://doi.org/10.3389/fpsyg.2016.01093

Schoeller, F., Perlovsky, L., & Arseniev, D. (2018). Physics of mind: Experimental confirmations of theoretical predictions. *Physics of Life Reviews*, *25*, 45–68. https://doi.org/10.1016/j.plrev.2017.11.021

Schreiber, J. (2018). pomegranate: Fast and flexible probabilistic modeling in Python. *Journal of Machine Learning Research*, *18*(164), 1–6.

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology*, *4*, 837. https://doi.org/10.3389/fpsyg.2013.00837

Schubert, E., North, A. C., & Hargreaves, D. J. (2016). Aesthetic experience explained by the affect-space framework. *Empirical Musicology Review*, *11*(3–4), 330–345. https://doi.org/10.18061/emr.v11i3-4.5115

Schubert, T. W., Zickfeld, J. H., Seibt, B., & Fiske, A. P. (2018). Moment-to-moment changes in feeling moved match changes in closeness, tears, goosebumps, and warmth: Time series analyses. *Cognition and Emotion*, *32*(1), 174–184. https://doi.org/10.1080/02699931.2016.1268998

Schurtz, D. R., Blincoe, S., Smith, R. H., Powell, C. A. J., Combs, D. J. Y., & Kim, S. H. (2012). Exploring the social aspects of goose bumps and their role in awe and envy. *Motivation and Emotion*, *36*, 205–217. https://doi.org/10.1007/s11031-011-9243-8

Seibt, B., Schubert, T. W., Zickfeld, J. H., & Fiske, A. P. (2017). Interpersonal closeness and morality predict feelings of being moved. *Emotion*, *17*(3), 389–394. https://doi.org/10.1037/emo0000271

Seibt, B., Schubert, T. W., Zickfeld, J. H., Zhu, L., Arriaga, P., Simão, C., Nussinson, R., & Fiske, A. P. (2018). Kama muta: Similar emotional responses to touching videos across the United States, Norway, China, Israel, and Portugal. *Journal of Cross-Cultural Psychology*, *49*(3), 418–435. https://doi.org/10.1177/0022022117746240

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Silvia, P. J., Fayn, K., Nusbaum, E. C., & Beaty, R. E. (2015). Openness to experience and awe in response to nature and music: Personality and profound aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(4), 376–384. https://doi.org/10.1037/aca0000028

Silvia, P. J., & Nusbaum, E. C. (2011). On personality and piloerection: Individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(3), 208–214. https://doi.org/10.1037/a0021914

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, *19*(2), 110–120. https://doi.org/10.1177/0305735691192002

Smith, J. D., & Melara, R. J. (1990). Aesthetic preference and syntactic prototypicality in music: 'Tis the gift to be simple. *Cognition*, *34*(3), 279–298. https://doi.org/10.1016/0010-0277(90)90007-7

Solberg, R. T., & Dibben, N. (2019). Peak experiences with electronic dance music: Subjective experiences, physiological responses, and musical characteristics of the break routine. *Music Perception*, *36*(4), 371–389. https://doi.org/10.1525/mp.2019.36.4.371

Starcke, K., von Georgi, R., Tiihonen, T. M., Laczika, K.-F., & Reuter, C. (2019). Don't drink and chill: Effects of alcohol on subjective and physiological reactions during music listening and their relationships with personality and listening habits. *International Journal of Psychophysiology*, *142*, 25–32. https://doi.org/10.1016/j.ijpsycho.2019.06.001

Stark, E. A., Vuust, P., & Kringelbach, M. L. (2018). Music, dance, and other art forms: New insights into the links between hedonia (pleasure) and eudaimonia (well-being). In J. F. Christensen & A. Gomila (Eds.), *The arts and the brain: Psychology and physiology beyond pleasure* (pp. 129–152). Academic Press. https://doi.org/10.1016/bs.pbr.2018.03.019

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Strick, M., de Bruin, H. L., de Ruiter, L. C., & Jonkers, W. (2015). Striking the right chord: Moving music increases psychological transportation and behavioral intentions. *Journal of Experimental Psychology: Applied*, *21*(1), 57–72. https://doi.org/10.1037/xap0000034

Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*(2), 213–226. https://doi.org/10.1080/09524622.2008.9753600

Sumpf, M., Jentschke, S., & Koelsch, S. (2015). Effects of aesthetic chills on a cardiac signature of emotionality. *PLOS ONE*, *10*(6), e0130117. https://doi.org/10.1371/journal.pone.0130117

Sutherland, M. E., Grewe, O., Egermann, H., Nagel, F., Kopiez, R., & Altenmüller, E. (2009). The influence of social situations on music listening. *Annals of the New York Academy of Sciences*, *1169*, 363–367. https://doi.org/10.1111/j.1749-6632.2009.04764.x

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *Journal of Neuroscience*, *31*(10), 3805–3812. https://doi.org/10.1523/jneurosci.5561-10.2011

Thompson, C., Parry, J., Phipps, D., & Wolff, T. (2020). *spotifyr: R wrapper for the 'Spotify' Web API (R package version 2.1.1.)* https://github.com/charlie86/spotifyr

Tihanyi, B. T. (2016). A zenei bizsergés pszichofiziológiai háttere és terápiás felhasználása. *Mentálhigiéné és Pszichoszomatika*, *17*(1), 19–36. https://doi.org/10.1556/0406.17.2016.1.2

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal*, *146*, 10.

Trainor, L. J., & Zatorre, R. J. (2016). The neurobiology of musical expectations from perception to emotion. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The oxford handbook of music psychology* (2nd, pp. 285–305). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.21

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Verhaeghen, P. (2018). Once more, with feeling: The role of familiarity in the aesthetic response. *The Psychological Record*, *68*, 379–384. https://doi.org/10.1007/s40732-018-0312-1

Vitz, P. C. (1966). Affect as a function of stimulus variation. *Journal of Experimental Psychology*, *71*(1), 74–79. https://doi.org/10.1037/h0022619

Vuoskoski, J. K., & Eerola, T. (2017). The pleasure evoked by sad music is mediated by feelings of being moved. *Frontiers in Psychology*, *8*, 439. https://doi.org/10.3389/fpsyg.2017.00439

Vuoskoski, J. K., Zickfeld, J. H., Alluri, V., Moorthigari, V., & Seibt, B. (2022). Feeling moved by music: Investigating continuous ratings and acoustic correlates. *PLOS ONE*, *17*(1), e0261151. https://doi.org/10.1371/journal.pone.0261151

Vuust, P., & Kringelbach, M. L. (2010). The pleasure of music. In M. L. Kringelbach & K. C. Berridge (Eds.), *Pleasures of the brain* (pp. 255–269). Oxford University Press.

Wang, C., Benetos, E., Meng, X., & Chew, E. (2019). HMM-based glissando detection for recordings of chinese bamboo flute. In I. Barbancho, L. J. Tardón, A. Peinado, & A. M. Barbancho (Eds.), *SMC 2019 Proceedings of the 16th Sound & Music Computing Conference* (pp. 545–550). Universidad de Málaga.

Warrenburg, L. A. (2020). Choosing the right tune: A review of music stimuli used in emotion research. *Music Perception*, *37*(3), 240–258. https://doi.org/10.1525/mp.2020.37.3.240

Wassiliwizky, E., Jacobsen, T., Heinrich, J., Schneiderbauer, M., & Menninghaus, W. (2017). Tears falling on goosebumps: Co-occurrence of lacrimation and emotional piloerection indicates a psychophysiological climax in emotional arousal. *Frontiers in Psychology*, *8*, 41. https://doi.org/10.3389/fpsyg.2017.00041

Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The emotional power of poetry: Neural circuitry, psychophysiology and compositional principles. *Social Cognitive and Affective Neuroscience*, *12*(8), 1129–1240. https://doi.org/10.1093/scan/nsx069

Wassiliwizky, E., Wagner, V., Jacobsen, T., & Menninghaus, W. (2015). Art-elicited chills indicate states of being moved. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(4), 405–416. https://doi.org/10.1037/aca0000023

Weth, K., Raab, M. H., & Carbon, C.-C. (2015). Investigating emotional responses to self-selected sad music via self-report and automated facial analysis. *Musicae Scientiae*, *19*(4), 412–432. https://doi.org/10.1177/1029864915606796

Williams, P. G., Johnson, K. T., Curtis, B. J., King, J. B., & Anderson, J. S. (2018). Individual differences in aesthetic engagement are reflected in resting-state fMRI connectivity: Implications for stress resilience. *NeuroImage*, *179*, 156–165. https://doi.org/10.1016/j.neuroimage.2018.06.042

Wundt, W. (1863). *Vorlesungen über die Menschen- und Tierseele*. Leopold Voss.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zatorre, R. J. (2003). Music and the brain. *Annals of the New York Academy of Sciences*, *999*(1), 4–14. https://doi.org/10.1196/annals.1284.001

Zatorre, R. J. (2015). Musical pleasure and reward: Mechanisms and dysfunction. *Annals of the New York Academy of Sciences*, *1337*(1), 202–211. https://doi.org/10.1111/nyas.12677

Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(Suppl. 2), 10430–10437. https://doi.org/10.1073/pnas.1301228110

Zentner, M., & Eerola, T. (2010). Self-report measures and models. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 187–221). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0008

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. https://doi.org/10.1037/1528-3542.8.4.494

Zickfeld, J. H., Schubert, T. W., Seibt, B., Blomster, J. K., Arriaga, P., Basabe, N., Blaut, A., Caballero, A., Carrera, P., Dalğar, I., Ding, Y., Dumont, K., Gaulhofer, V., Gračanin, A., Gyenis, R., Hu, C.-P., Kardum, I., Lazarević, L., Mathew, L., … Fiske, A. P. (2019). Kama muta: Conceptualizing and measuring the experience often labelled being moved across 19 nations and 15 languages. *Emotion*, *19*(3), 402–424. https://doi.org/10.1037/emo0000450

Zickfeld, J. H., Schubert, T. W., Seibt, B., & Fiske, A. P. (2017). Empathic concern is part of a more general communal emotion. *Frontiers in Psychology*, *8*, 723. https://doi.org/10.3389/fpsyg.2017.00723

Zickfeld, J. H., Schubert, T. W., Seibt, B., & Fiske, A. P. (2019). Moving through the literature: What is the emotion often denoted being moved? *Emotion Review*, *11*(2), 123–139. https://doi.org/10.1177/1754073918820126