

Reinhard Kopiez (Hrsg.)



Musikpsychologie –
Empirische Forschungen –
Ästhetische Experimente

Band 33

Jahrbuch der Deutschen Gesellschaft
für Musikpsychologie

Jahrbuch der Deutschen Gesellschaft für Musikpsychologie

Band 33

Herausgeber der Reihe:
Deutsche Gesellschaft für
Musikpsychologie (DGM)

Herausgeber des Bandes (Editor-in-Chief):
Reinhard Kopiez

Redaktion:
Nina Düvel, Kilian Vogt

Wissenschaftlicher Beirat (Editorial Board):

Anna Rita Addressi (Bologna, Italien)	Steffen Lepa (Berlin)
Ralf von Appen (Wien, Österreich)	Tim Loepthien (München)
Claudia Bullerjahn (Gießen)	Kai Lothwesen (Trossingen)
Veronika Busch (Bremen)	Christoph Louven (Osnabrück)
Claus-Christian Carbon (Bamberg)	Adina Mornell (München)
Franziska Degé (Frankfurt a.M.)	Daniel Müllensiefen (London, UK)
Tuomas Eerola (Durham, UK)	Michael Oehler (Osnabrück)
Hauke Egermann (Köln)	Friedrich Platz (Stuttgart)
Timo Fischinger (Frankfurt a.M.)	Christoph Reuter (Wien, Österreich)
Klaus Frieler (Frankfurt a.M.)	Nicolas Ruth (München)
Werner Goebel (Wien, Österreich)	Stephan Sallat (Halle)
Johannes Hasselhorn (Erlangen)	Thomas Schäfer (Berlin)
Jan-Peter Herbst (Huddersfield, UK)	Holger Schramm (Würzburg)
Ann-Kristin Herget (Dortmund)	Emery Schubert (Sydney, Australien)
Erkki Huovinen (Stockholm, Schweden)	Eva Schurig (Oldenburg)
Jens Knigge (Levanger, Norwegen)	Olivier Senn (Luzern, Schweiz)
Gunter Kreutz (Oldenburg)	Kai Siedenburg (Oldenburg)
Elke Lange (Frankfurt a.M.)	Felix Thiesen (Hannover)
André Lee (Hannover)	Renee Timmers (Sheffield, UK)
Andreas C. Lehmann (Würzburg)	Clemens Wöllner (Freiburg)

Für den DGM-Vorstand:

Ann-Kristin Herget (Hannover)	Felix Thiesen (Dortmund)
Jesper Hohagen (Freiburg)	Clemens Wöllner (Freiburg)
Jörg Mühlhans (Wien, Österreich)	Anna Wolf (Weimar)
Kathrin Schlemmer (Eichstätt)	

Das Jahrbuch Musikpsychologie publiziert Open Access First unter
<http://jbdgm.psychopen.eu>

Reinhard Kopiez (Hrsg.)

Musikpsychologie –
Empirische Forschungen –
Ästhetische Experimente



Waxmann 2026
Münster • New York

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

Jahrbuch der Deutschen Gesellschaft für Musikpsychologie, Band 33

ISSN 2511-8277

Print-ISBN 978-3-8188-0151-9

E-Book-ISBN 978-3-8188-5151-4

© Waxmann Verlag GmbH, 2026
Steinfurter Str. 555, 48159 Münster

www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Pleßmann Design, Ascheberg
Satz: Roger Stoddart, Münster

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des
Verlages in irgendeiner Form reproduziert oder unter Verwendung
elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Inhalt

Editorial

Reinhard Kopiez

Editorial	9
-----------------	---

Forschungsberichte

Hanna Mütze, Friedrich Platz, Veronika Busch

The Acquisition of Tonal Hierarchies in Western Music During School Years: A Re-Analysis of 40 Years of Research	15
---	----

Der Erwerb tonaler Hierarchien in westlicher Musik während der Schulzeit:
eine Re-Analyse der Ergebnisse aus 40 Jahren Forschung

Kim Ehler, Nina Düvel, Kai Lothwesen

Prävalenz und Prädiktoren von Neuromythen bei (Musik-)Studierenden: eine Teilreplikation der Studie von Düvel et al. (2017)	49
--	----

Prevalence and Predictors of Neuromyths in (Music) Students:
A Partial Replication of the Study by Düvel et al. (2017)

Steffen Lepa, Luzie Ahrens, Martin Pfeleiderer

Dimensionen der Groove-Affordanz (DGA) beim Hören populärer Musik: ein neues Messinstrument und eine vergleichende Validierungsstudie	75
--	----

Dimensions of Groove Affordances (DGA) When Listening to Popular Music:
A New Measurement Instrument and a Comparative Validation Study

Nicholas Meier, Kilian Sander, Anton Schreiber, Reinhard Kopiez

The Creative Musical Achievement of AI Systems Compared to Music Students: A Replication of the Study by Schreiber et al. (2024)	113
---	-----

Die kreativen musikalischen Leistungen von KI-Systemen im Vergleich zu
Musikstudierenden: eine Replikation der Studie von Schreiber et al. (2024)

Edoardo Passarotto, Patrizia Bieber, Daniel Müllensiefen

Adaptation and Validation of the Deliberate Practice in Music Inventory for Child Musicians	135
--	-----

Adaptierung und Validierung des Deliberate Practice in Music Inventory für Kinder

Nina Düvel, Franziska Altemeier

Open Science Practices and New Statistics in the Yearbook of Music Psychology: A Bibliometric Analysis	159
---	-----

Verbreitung von Open-Science-Praktiken und New Statistics im Jahrbuch
Musikpsychologie: eine bibliometrische Untersuchung

Berichte

Lilian Peters

Art in Motion 2025: The Drive to Thrive, 16.–18. Mai 2025,
Hochschule für Musik und Theater München 203

Miguel Machulla

Digitalisierung in der Musikpsychologie – 40. Jahrestagung der
Deutschen Gesellschaft für Musikpsychologie (DGM), 6.–8. September 2024,
Hochschule für Musik und Theater München 206

Editorial

Editorial

Reinhard Kopiez¹

[1] Hanover Music Lab, Hochschule für Musik, Theater und Medien Hannover

Zusammenfassung

Im Editorial erläutert der Herausgeber die Hintergrundprozesse der Arbeit für das *Jahrbuch Musikpsychologie* und dankt allen Mitwirkenden. Überlegungen zu zukünftigen Aufgaben werden diskutiert.

Abstract

In the Editorial, the Editor-in-Chief explains the background processes during work for the *Yearbook of Music Psychology* and thanks all contributors. Considerations for future tasks are discussed.

Liebe Leser*innen,

hiermit liegt der dritte Band des *Jahrbuchs Musikpsychologie* (JBDGM) während meiner Amtszeit als Herausgeber vor, welcher auch der letzte in Buchform sein wird. Auf Vorschlag des Vorstands der DGM und mit Beschluss durch die Mitgliederversammlung wird das JBDGM zukünftig nur noch als reine Online-Publikation erscheinen (weiterhin unter <https://jbdgm.psychopen.eu/index.php/jbdgm>). Auch in diesem letzten Band in Print gibt es wieder eine erfreuliche Anzahl hochkarätiger Beiträge und Berichte.

So eine umfangreiche Publikation wird durch eine Vielzahl von Personen unterstützt, die kontinuierlich für die Zeitschrift arbeiten: Da wären die beiden Redaktionsassistent*innen Nina Düvel und Kilian Vogt, durch deren kritischen Blick alle Manuskripte von Anfang an formal geprüft und in Absprache mit den Autor*innen in Hinblick auf die einzuhaltenden Standards optimiert werden.

Mein Dank gilt auch dem Editorial Board, dessen Mitglieder das JBDGM mit Reviews der Artikel entscheidend unterstützen. Seit 2025 sind Anna Rita Addressi (University of Bologna, Italien), Tuomas Eerola (Durham University, UK), Jesper Hohagen (Hochschule für Musik Freiburg/Universitätsklinikum Freiburg), Emery Schubert (University of New South Wales, Sydney, Australien) und Renee Timmers (University of Sheffield, UK) neu hinzugekommen. Sie werden uns mit ihrer Expertise in weiteren Bereichen der empirischen Musikforschung unterstützen. Die Webseite des JBDGM zum Editorial Team gibt hier weiterführende Informationen (https://jbdgm.psychopen.eu/index.php/JBDGM/editorial_team).

Ich möchte mich auch beim Copy-Editing-Team der Plattform PsychOpen für die administrative Betreuung der Open-Access-Publikation des JBDGM bedanken. Ohne die besondere Unterstützung durch Gerrit Fröhlich sowie Marina Kuhn und die anderen Mitglieder des Teams wäre diese professionelle Veröffentlichungsform nicht möglich. Auch die Zusammenarbeit mit Alexandra Wilken vom Waxmann Verlag erwies sich wieder als sehr konstruktiv bei der Herstellung der Druckversion des JBDGM.

Eine peer-reviewte Zeitschrift lebt immer auch vom Einsatz der Gutachtenden, und so möchte ich folgenden Personen für ihren Einsatz bei der Begutachtung der für Band 33 des JBDGM eingereichten Beiträge danken:

Toni Bechtold, Patrizia Bieber, Benjamin Burkhart, Theresa Demmer, Tuomas Eerola, Hauke Egermann, Klaus Frieler, Jan-Peter Herbst, Jens Knigge, Philippe Labonde, Hsin-Rui Lin, Jörg Mühlhans, Friedrich Platz, Stephan Sallat, Thomas Schäfer, Holger Schramm, Eva Schurig und Wolf-Georg Zaddach.

Letztlich möchte ich auch Ihnen als einreichende Autor*innen danken, denn von Ihren Forschungs- und Publikationsaktivitäten lebt diese Zeitschrift.

Das JBDGM steht im Feld der wissenschaftlichen Zeitschriften im Wettbewerb mit zahlreichen anderen qualitativ hochwertigen Publikationsmöglichkeiten, doch bin ich davon überzeugt, dass es ein hohes Potential für die Scientific Community im internationalen Vergleich besitzt. In Anlehnung und Weiterentwicklung der bereits von Auhagen et al. (2018) aufgestellten Leistungs- und Alleinstellungsmerkmale wären hervorzuheben:

- **Hoher Qualitätsanspruch:** Eingereichte Forschungsberichte durchlaufen ein anonymisiertes, online-basiertes Peer-Review-Verfahren nach internationalen Standards. Neben dem Peer-Review durchlaufen die Beiträge eine Plagiatsprüfung.
- **Schnelles Publizieren:** Positiv begutachtete Beiträge werden fortlaufend jeweils zeitnah im JBDGM veröffentlicht. Es gibt keine Wartezeit auf Fertigstellung der Printversion. Von der Einreichung bis zur weltweit sichtbaren Publikation vergehen im günstigsten Fall nur wenige Wochen. Bei Band 33 lag die Zeit bis zur ersten herausgeberischen Entscheidung im Median bei ca. 50 Tagen (deutlich unter dem empfohlenen Richtwert der APA von weniger als drei Monaten) und die Acceptance Rate betrug 72%. Sämtliche Beiträge erhalten eine Identifikationsnummer (DOI) und sind damit sofort voll zitierfähig.
- **Online Supplement:** Neben den eigentlichen Aufsätzen sollen im JBDGM im Rahmen einer Strategie für Open Data und Open Materials auch Anhänge wie experimentelle Materialien veröffentlicht werden (sofern die Rechte dafür vorliegen).
- **Internationalität:** Im JBDGM sind Publikationen in deutscher und englischer Sprache möglich. Jeder Beitrag enthält zudem jeweils eine Kurzfassung/ein Abstract in deutscher und englischer Sprache.
- **Sichtbarkeit:** Alle Beiträge werden in die einschlägigen musikwissenschaftlichen und psychologischen Datenbanken (RILM, PubPsych, Psynindex etc.) aufgenommen und sind sofort weltweit recherchierbar und nutzbar.
- **Open-Science-Praktiken:** Das JBDGM empfiehlt nachdrücklich die Veröffentlichung der zu einem Beitrag gehörenden Forschungsdaten und folgt damit den Leitlinien zur Guten Wissenschaftlichen Praxis u. a. der Deutschen Forschungsgemeinschaft (DFG) und der Deutschen Gesellschaft für Psychologie (DGP). Für die Veröffentlichung der Forschungsdaten wird dabei in erster Linie das Datenrepositorium PsychArchives des Leibniz-Instituts für Psychologie (ZPID; <https://leibniz-psychology.org>) genutzt. Im Zeitalter von Open Data ist eine nachhaltige Nut-

zung von Forschungsdaten (z. B. für Meta-Analysen) auch ein Gebot der Stunde und ein Zeichen für ein erfolgreiches Forschungsdatenmanagement. Mit dem vorliegenden Band 33 des JBDGM kennzeichnet das ZPID jeden Beitrag auch dahingehend, welche der vier Merkmale von folgenden Open-Sciences-Praktiken im jeweiligen Artikel umgesetzt wurden: Für die Merkmale Präregistrierung, Open Data, Open Materials und Open Code/Software erscheinen nun als Hinweis auf besondere Qualitätsmerkmale in der Artikelkopfzeile bis zu vier kleine Icons (die sogenannten Open Science Badges; Center for Open Science, 2026). Dies unterstützt die Musikpsychologie dabei, den Anschluss an die von der European Research Executive Agency (2026) geförderten internationalen Standards in der Open-Science-Praxis umzusetzen. Wie die bibliometrische Analyse von Düvel und Altemeier (2026, in diesem Band) für zwischen 2012 und 2024 im JBDGM veröffentlichte Artikel zeigt, gibt es bei diesen Praktiken für musikpsychologische Publikationen noch Entwicklungspotential.

- **Diamond Open Access:** Sowohl die Nutzung der Inhalte des JBDGM für die Leser*innen als auch die Einreichung, Begutachtung und Veröffentlichung von Beiträgen für die Autor*innen sind grundsätzlich kostenfrei. Dies unterscheidet die Publikationspraxis des JBDGM von den meisten anderen (und häufig kommerziellen) Open-Access-Zeitschriften und wird als Diamond Open Access bezeichnet (Open Access Network, 2026). Die Plattform *PsychOpen* des ZPID bietet hierbei dem JBDGM hervorragende Voraussetzungen für die Umsetzung eines nicht-kommerziellen Open-Access-Ansatzes in der psychologischen Forschung.

Als wichtiger Schritt für die Internationalisierung des JBDGM wurde im Juli 2025 der Antrag auf Aufnahme in den *Scopus Cite Score* des Elsevier-Verlags gestellt. Zuvor war bedauerlicherweise der Antrag auf Aufnahme des JBDGM in das Web of Science und damit die Möglichkeit zur Berechnung des *Journal Impact Factors* im ersten Anlauf abgelehnt worden. Hauptgrund dafür war die aus Sicht der Evaluationskommission zu unregelmäßige Erscheinungsweise (Timeliness and/or Publication Volume), da es aufgrund einer zu geringen Anzahl an Artikeln für Band 32 (2023/2024) einen Doppelband gab. Der nun neu angelaufene Evaluationsprozess für den *Scopus Cite Score* dauert bis zu einem Jahr. Hierdurch wird mittelfristig die Berechnung eines dem *Journal Impact Factor* vergleichbaren bibliometrischen Indexes für das JBDGM ermöglicht. Dies wird die internationale Sichtbarkeit des JBDGM weiter erhöhen. Allerdings ist eine positive Evaluierung an bestimmte Kriterien gebunden, wozu zentral die Regelmäßigkeit der Beitragspublikation gehört (Timeliness of Publication). Die kontinuierliche Einreichung hochwertiger Manuskripte (in Englisch oder Deutsch) ist demnach eine der wichtigen Voraussetzungen für eine positive Evaluation. Deshalb meine Bitte: Denken Sie auch in Zukunft an das JBDGM als interessantes Outlet für Ihre Manuskripte und bewerben Sie es auch im Kreis Ihrer Kolleg*innen, Doktorand*innen und Master-Kandidat*innen!

Nun wünsche ich eine anregende Lektüre beim *Jahrbuch Musikpsychologie* und freue mich auch in Zukunft über zahlreiche Forschungsbeiträge für das JBDGM.

Reinhard Kopiez

Herausgeber des *Jahrbuch Musikpsychologie*

(Editor-in-Chief)

Literatur

Auhagen, W., Bullerjahn, C., & Louven, C. (2018). Das neue ‚Open Access First‘-Publikationsmodell des Jahrbuchs Musikpsychologie: Geleitwort der Herausgeber. *Jahrbuch Musikpsychologie*, 28, 11–13. <https://www.doi.org/10.5964/jbdgm.2018v28.22>

Center for Open Science (15. Januar 2026). *Open Science Badges enhance openness, a core value of scientific practice*. <https://www.cos.io/initiatives/badges>

Düvel, N., & Altemeier, F. (2026). Open Science Practices and New Statistics in the Yearbook of Music Psychology: A Bibliometric Analysis. *Jahrbuch Musikpsychologie*, 33, 159–199. <https://doi.org/10.5964/jbdgm.229>.

European Research Executive Agency (15. Januar 2026). *Open Science*. https://rea.ec.europa.eu/open-science_en.

Open Access Network (15. Januar 2026). *Grün, Gold und Diamond*. <https://open-access.network/informieren/open-access-grundlagen/open-access-gruen-und-gold#c17391>

Forschungsberichte

The Acquisition of Tonal Hierarchies in Western Music During School Years: A Re-Analysis of 40 Years of Research

Der Erwerb tonaler Hierarchien in westlicher Musik während der Schulzeit: eine Re-Analyse der Ergebnisse aus 40 Jahren Forschung

Hanna Mütze¹ , Friedrich Platz² , Veronika Busch¹ 

[1] Department of Musicology and Music Education, University of Bremen, Bremen, Germany.

[2] Institut für Ästhetisch-Kulturelle Wissenschaft und Praxis, Europa-Universität Flensburg, Flensburg, Germany.

Jahrbuch Musikpsychologie, 2025, Vol. 33, Article e205, <https://doi.org/10.5964/jbdgm.205>

Received: 2024-08-26 • Accepted: 2025-09-22 • Published (VoR): 2025-11-12.

Reviewed by: Reinhard Kopiez; Klaus Frieler; Thomas Schäfer.

Corresponding Author: Hanna Mütze, Department of Musicology and Music Education, University of Bremen, Bibliothekstraße 1, 28359 Bremen, Germany. E-mail: muetzeh@uni-bremen.de



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.



Supplementary Materials: Code, Data, Materials [see Index of Supplementary Materials]

Abstract

Understanding the relationships between different pitches as a form of tonality is a key element of listening skills in Western tonal music. Tonal hierarchies (i.e., genre-dependent differing prominence of tones) are reflected in the internal representations of tonal hierarchies (IRTH) in long-term memory. Over the past 40 years, research on how individuals – primarily students aged 6 to 15, as well as adults – acquire IRTH has yielded varied and sometimes contradictory conclusions about the timeline and underlying mechanisms of this process. This review aims to synthesize the evidence and critically examine potential reasons for the heterogeneity in prior findings. To this end, two approaches were applied. First, a Bayesian three-level meta-analysis of 60 effect sizes from 16 studies, reported in 13 articles, revealed a medium difference in IRTH sensitivity between younger and older participants. Second, a model comparison analysis based on cross-sectional data from a single study revealed a non-linear growth dynamic, with a larger increase during adolescence as the best model solution to describe the relationship between sensitivity and age. We also examined the considerable heterogeneity observed within and between studies, particularly how task-specific features of the operationalizations might account for these differences. These findings contribute to the development of theoretical models of music-related skill acquisition and suggest directions for future research.

Keywords: tonal development, probe tone, Bayesian meta-analysis, musical skill acquisition

Zusammenfassung

Das Verständnis der Beziehungen zwischen verschiedenen Tonhöhen als eine Form der Tonalität ist ein zentraler Bestandteil der Hörfähigkeit in westlich-tonaler Musik. Tonale Hierarchien (d. h. die genreabhängig unterschiedliche Prominenz von Tönen) spiegeln sich in internen Repräsentationen tonaler Hierarchien (IRTH) im Langzeitgedächtnis wider. In den vergangenen 40 Jahren hat die Forschung darüber, wie Teilnehmende – hauptsächlich Schulkinder im Alter von 6 bis 15 Jahren, aber auch Erwachsene – IRTH erwerben, unterschiedliche und mitunter widersprüchliche Schlussfolgerungen hinsichtlich des zeitlichen Ablaufs und der zugrunde liegenden Mechanismen hervorgebracht. In diesem Beitrag sollen die Erkenntnisse zusammengefasst und mögliche Gründe für die Heterogenität der Ergebnisse kritisch untersucht werden. Zu diesem Zweck wur-

den zwei Ansätze angewandt. Erstens ergab eine Bayesianische Meta-Analyse von 60 Effektgrößen aus 16 Studien, die in 13 Artikeln berichtet wurden, einen Medianunterschied der IRTH-Empfindlichkeit zwischen jüngeren und älteren Kindern im Umfang einer mittleren Effektgröße. Zweitens zeigte eine Modellvergleichsanalyse auf der Grundlage von Querschnittsdaten aus einer einzigen Studie eine nichtlineare Wachstumsdynamik mit einem größeren Anstieg während der Adoleszenz und nicht im Kindes-/Grundschulalter als beste Modelllösung zur Beschreibung der Beziehung zwischen Sensibilität und Alter. Eine Heterogenitätsanalyse zeigte abschließend, dass sich die Verteilung der Effektstärken früherer Studien maßgeblich auf aufgabenspezifische Merkmale im Zuge der Operationalisierungen der Repräsentation tonaler Hierarchien zurückführen und erklären lassen. Diese Ergebnisse tragen zur Entwicklung theoretischer Modelle über den Erwerb musikbezogener Fertigkeiten bei und eröffnen Perspektiven für zukünftige Forschungen.

Schlüsselwörter: tonale Entwicklung, Prüfverfahren, bayessche Metaanalyse, musikalischer Fertigkeitserwerb

Tonality is reflected by the variations in prominence given to tones on a scale (Piston, 1978; von Helmholtz, 1896) because of differences in their frequency of occurrence, metrical position, and duration (Prince & Schmuckler, 2014; Verosky, 2021). Most Western music considers the concept of tonality as:

... one of the main conceptual categories of Western musical thought, [referring] to the orientation of melodies and harmonies towards a referential (or tonic) pitch class. In the broadest possible sense, however, it refers to systematic arrangements of pitch phenomena and relations between them. (Hyer, 2021, p. 1)

Several music psychology models suggest that tonality perception is linked to the varying stabilities perceived in individual tones (Schmuckler, 2016). Krumhansl and Shepard (1979) demonstrate that tonal hierarchy is reflected in individuals' stability ratings of tones, depending on the hierarchical level in the preceding tonal context. This hierarchy in classical/romantic music includes four levels: tonic, tonic triad, other diatonic tones, and non-diatonic tones (Krumhansl & Kessler, 1982). Krumhansl and Keil (1982) conclude that tonal hierarchy is stored in long-term memory as internal representations of tonal hierarchies (IRTH). In this paper, the ability to perceive the different stabilities of tones is referred to as sensitivity to IRTH.

Based on the above, our analyses focus on the IRTH acquisition processes (i.e., how IRTH evolves with increasing age). The scientific debate about the timing and mechanisms of IRTH development is ongoing (Patel, 2021), with narrative summaries (e.g., Corrigan & Schellenberg, 2016; Gembris, 2017a) and theoretical models offering varied and sometimes contradictory conclusions. Earlier theories, such as Brehmer's concept of "tonal giftedness" (German: "Begabung"), suggest that this ability is fully developed by age 7 (Kühn, n.d., cited in Brehmer, 1925, p. 172). Gordon (2012) argues that musical aptitude and audiation show significant development before stabilizing around age 9, with limited improvement thereafter. This extended developmental trajectory contrasts with Brehmer's earlier views.

Gardner's (1973) theory includes two stages of artistic development concluding around age 7, whereas Swanwick and Tillman (1986) propose four stages extending into adolescence. Hargreaves (1996) supports the latter view, emphasizing the impor-

tance of ages 8 to 15 for mastering musical rules. Conversely, Serafine (1988) suggests a more restricted sensitive period for tonal learning from ages 8 to 10. In stark contrast, recent research highlights statistical learning as a key factor in lifelong IRTH development (Jonaitis & Saffran, 2009; Vuvan, 2013), with evidence showing that this process begins as early as infancy (Saffran et al., 1999). Thus, whether IRTH has already developed fully by one's school years or continues to develop during that time remains an open question.

Regarding primary studies, Krumhansl and Keil (1982) identify a clear age-related progression in IRTH: first- and second-graders prefer diatonic tones over non-diatonic tones, third- and fourth-graders prefer the tonic triad over other diatonic tones, and fifth-graders as well as adults prefer the tonic overall. Maier-Karius and Schwarzer (2011), Paananen (2007, 2009), and Schwarzer et al. (1993) replicate this trajectory; however, Matsunaga et al. (2020), Schellenberg et al. (2005), Speer and Meeks (1985), and Wilson and Wales (1995) find either earlier or later IRTH acquisition.

The failure to replicate previous findings may be related to several factors, such as participant characteristics or methodological differences. One factor might be formal musical training, defined as systematic instruction in music theory or practice (Hanna-Pladdy & MacKay, 2011). Although some studies suggest that formal training enhances IRTH acquisition (Corrigall et al., 2022; Mandikal Vasuki et al., 2016), others find no such effect, primarily attributing acquisition to implicit learning processes (Cui, 2019; Müllensiefen et al., 2014). Moreover, the influence of formal training may depend on factors such as age, ongoing training, and varying types of operationalization (Asztalos & Csapo, 2017; Müllensiefen et al., 2022; Zhang et al., 2020). Cultural background (Matsunaga et al., 2020), gender (Lin, 2023), and socioeconomic status (Miles et al., 2016) may also modulate IRTH acquisition.

Studies may also suffer from a lack of statistical power, which increases the likelihood of failing to detect a true population effect (Ellis, 2010). Finally, the specific method used to measure IRTH may affect the results. In contrast to the predictions of earlier theories on the development of tonal hierarchies, which rely on explicit measures with (partially) time-independent response formats, several studies using implicit measures report detecting tonal knowledge even in early childhood, indicating an earlier onset of this ability than previously suggested (Jentschke et al., 2005; Politimou et al., 2021; Trehub et al., 1999). Corrigall et al. (2022) find that task characteristics influence IRTH acquisition in school-aged children, with significant differences only emerging in explicit tasks.

In summary, despite 40 years of vibrant research, we are faced with a highly heterogeneous body of literature, making it difficult to draw clear evidence on the shape of IRTH skill acquisition. This complexity pertains not only to identifying the general developmental trajectory but also to distinguishing between specific phases of skill acquisition, such as the phase of initial skill acquisition and the potential onset of a saturation effect – a plateau in skill improvement in relation to the conditions (e.g., formal instruction) under which differences in the shapes and effects of skill acquisition occur.

Aims and Objectives

We conducted a critical review of four decades of research on the development of IRTH sensitivity. Despite a sizeable volume of studies, the findings remain heterogeneous and partially contradictory, particularly regarding developmental trajectories and the influence of musical training. Building on Gordon's (2012) theory of music learning, which posits that core aspects of tonal sensitivity are largely developed by age 9, we formulated four exploratory research questions. By examining these questions, this study aimed to (1) assess the magnitude of age-related changes in IRTH sensitivity, (2) examine whether sensitivity increases significantly beyond age 9, (3) investigate the role of musical training, and (4) explore the effects of different operationalizations of IRTH measurement.

Research Questions

Research Question 1 (RQ1). How substantial is the average age-related development of IRTH sensitivity?

Research Question 2 (RQ2). Does IRTH sensitivity continue to increase beyond the age of 9, or does it level off, as predicted by Gordon's learning theory?

Research Question 3 (RQ3). Is IRTH sensitivity significantly higher in musically trained individuals than in those without musical training?

Research Question 4 (RQ4). Does the implicit measurement of IRTH sensitivity via response time yield lower estimates than more explicit operationalizations?

Method

Procedure

We conducted a systematic and comprehensive literature search (What Works Clearinghouse, 2020) between January and April 2022 to identify eligible studies. The study selection followed a predefined process outlined in the review protocol (see Supplementary Material S1), which was not published before the review was conducted. In the final sample, we included only primary research studies of the preliminary literature corpus that used a hypothesis-testing approach, followed either a cross-sectional or longitudinal design, were published between 1982¹ and April 2022, investigated a sample of healthy elementary and secondary school students, and defined the measurement of IRTH sensitivity as a dependent variable based on either a listening task (probe tone paradigm, response time measures, and goodness-of-fit ratings in syntax-violation paradigms) or a creative production task (harmonization of a given mel-

1 We included only articles published after 1982 following Krumhansl and Kessler's (1982) pioneering work in this field.

ody, tonal composition, and improvisation to a given chord sequence). These operationalizations were further specified as described below.

First, the probe tone paradigm (Krumhansl & Shepard, 1979) involves presenting one or two of the 12 chromatic tones (the “probe”) after establishing the tonal context (e.g., an ascending major scale), followed by a brief silence. Participants rate how well the probe tones fit the tonal context. This process continues until all 12 chromatic tones are rated. The perceived stability of each tone is measured based on participants’ goodness-of-fit ratings. The correlation of these ratings with the prototypical tonal hierarchy (see above) defines participants’ sensitivity to IRTH. Although no standardized protocols or psychometric criteria² have been established, the probe tone paradigm has been consistently replicated across various contexts, demonstrating its robustness (Morgan et al., 2019; Sauvé et al., 2021).

Second, some studies use goodness-of-fit ratings in a syntax-violation paradigm (e.g., Corrigan et al., 2022). Participants judge how well the final tone of short melodies fits within the preceding harmonic context with varying degrees of tonal congruence (e.g., a tonic note such as “C” in C major versus a non-diatonic note such as “C#”). A greater difference in ratings between congruent and incongruent endings indicates IRTH acquisition, with higher ratings for congruent endings reflecting higher IRTH sensitivity.

Third, Schellenberg et al. (2005) and Corrigan et al. (2022) focus on response times (as introduced by Janata & Reisberg, 1988). Rather than explicitly evaluating the tonal congruence of melody endings, participants rate other musical features, such as the timbre of the final tone, while hearing a priming sequence with either a tonally congruent (expected) or incongruent (unexpected) final tone. Faster and more accurate responses are expected when the target tone or chord is more tonally congruent with the preceding context, indicating stronger IRTH sensitivity.

Fourth, other methods involve creative production tasks, such as composing (Wilson & Wales, 1995), improvising while listening to a pre-recorded tonal chord sequence (Paananen, 2003), or harmonizing a melody (Paananen, 2009). The outcome variables are defined by the degree of tonal fit and timing of the tones or chords used by the participants. Wilson and Wales (1995, p. 102) report a substantial to almost perfect inter-rater agreement (cf., Landis & Koch, 1977) for expert assessments of compositions ($.69 \leq \kappa \leq .92$).

In the next step, three independent and trained coders (including the first author) coded the included studies according to a previously developed protocol (see Supplemental Material S2). After calculating the initial inter-rater reliability, which showed almost perfect agreement ($\kappa = .92$), coding discrepancies were resolved by consensus. The studies vary in whether and how they report participants’ formal musical training; therefore, we calculated the percentage of those with training for each age group. Sufficient data are available in 11 studies, while other variables are coded as not applicable (“n/a”).

2 The lack of these psychometric criteria also applies to all of the measurement methods (exception: Wilson & Wales, 1995).

Data Analysis

Data were analyzed using two approaches. First, we conducted a Bayesian three-level meta-analysis to provide a quantitative summary of the systematic review, offer an overview, identify notable studies, and investigate differences in IRTH measurements. Second, we conducted a model comparison analysis of the cross-sectional data of a single study (Krumhansl & Keil, 1982) investigating the growth trajectory of IRTH acquisition.

In the meta-analytical approach, effect sizes were estimated from primary studies using the metafor package (Viechtbauer, 2010) in R (R Core Team, 2022) and, if needed, transformed into Cohen's d following Borenstein and Hedges (2019, pp. 214–234). The procedure for effect size calculations, raw data, and a markdown script for the Bayesian three-level meta-analysis are shown in Supplemental Materials S3, S4, and S5.

In the subsequent analysis step, the effect sizes were aggregated and weighted using a Bayesian three-level meta-analysis with the R package brms (Bürkner, 2017), following the procedure of Harrer et al. (2021). Unlike a one-level fixed effect model that only attributes variance to sampling error, the two-level random effects model separates variance into sampling errors³ and between-study heterogeneity (τ_1^2). This model captures inherent differences between studies, such as measurement methods.

Our model further includes a third level of variance (τ_2^2), reflecting the nested structure of effect sizes within studies and accounting for the variance within single studies, such as different age groups. This multilevel approach (for a formal model description, see Supplemental Material S6) better reflects the nested structure of our data and avoids the statistical issues of traditional univariate meta-analysis, which can misestimate heterogeneity and increase the risk of false positives (Harrer et al., 2021; Hedges, 2019). We compared models with differing levels of complexity by computing marginal likelihoods using the bridge sampling method implemented in the brms package.

We chose Bayesian meta-analysis for two key reasons. First, it allows for the explicit modeling of heterogeneity uncertainty (τ^2), which can yield more stable estimates in small-sample contexts – especially when applying weakly or moderately informative priors (Harrer et al., 2021). Notably, although Bayesian methods are not inherently robust to small samples or poor study quality, they provide a principled framework for incorporating prior knowledge and quantifying uncertainty. Second, prior information can improve the posterior estimation of key parameters, such as μ (intercept) and τ (standard deviations; Röver, 2020).

To address the context-dependence of choosing the prior parameter settings (Gelman et al., 2015), we initially used two separate and weakly informative zero-centered priors for μ and τ , as recommended by Williams et al. (2018) and Harrer et al. (2021, Setting Prior Distributions, para. 5):

3 As a consequence of Turner and Higgins (2019, p. 302), Bayesian models do not estimate the sampling error, as it is assumed to be known based on the sample size.