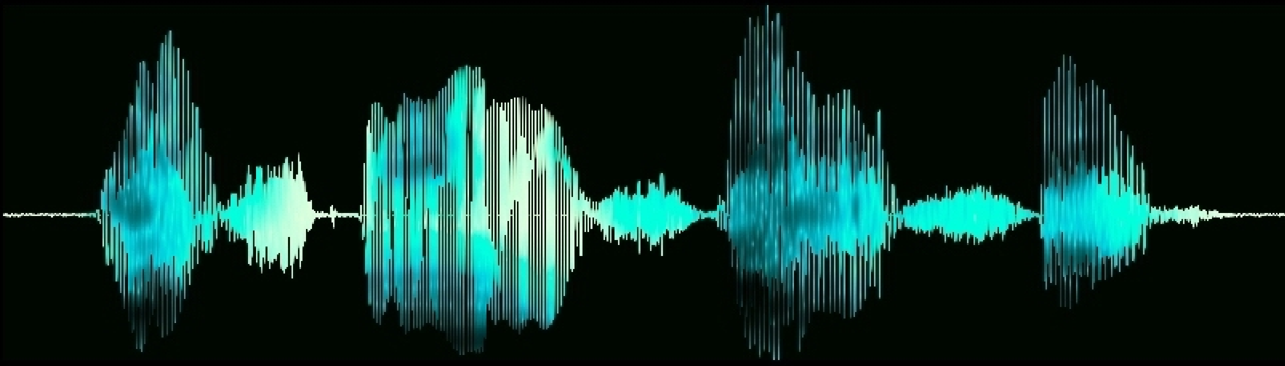
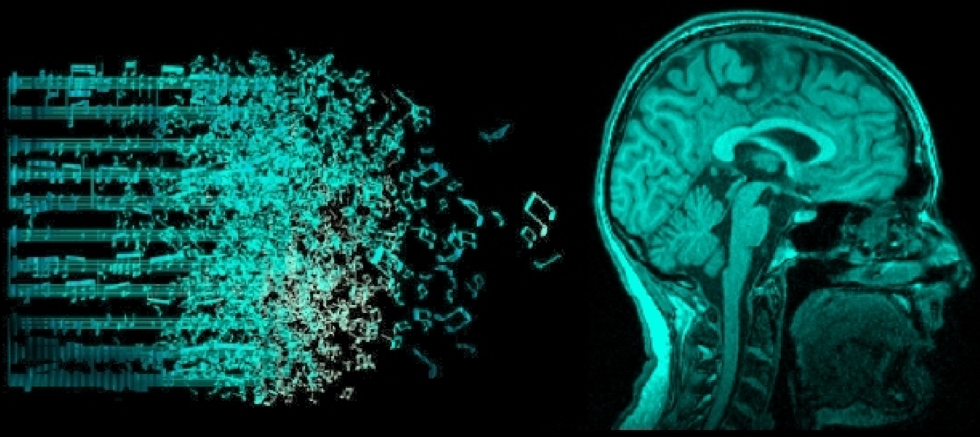


Untwisting Amusia

What behavior, brain waves and genetic underpinnings reveal about perception in congenital amusia



Jasmin Pfeifer

Untwisting Amusia

What behavior, brain waves and genetic underpinnings reveal about perception in congenital amusia

Jasmin Pfeifer

ISBN: 978-94-6421-648-6

NUR: 616

Author: Jasmin Pfeifer

Cover design: Jasmin Pfeifer

Printed by: Ipskamp Drukkers, Enschede, The Netherlands

© 2022 Jasmin Pfeifer. All rights reserved.

No part of this thesis may be reproduced or transmitted, in any form or by any means, without prior written permission of the author.

Untwisting Amusia

What behavior, brain waves and genetic underpinnings reveal about perception in congenital amusia

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 04 maart 2022, te 10:00 uur

door Jasmin Pfeifer

geboren te Meerbusch/Neuss

Promotiecommissie

Promotores:	prof. dr. P.P.G. Boersma	Universiteit van Amsterdam
	prof. dr. P. Indefrey	Heinrich-Heine-Universität Düsseldorf
Copromotores:	dr. S.R. Hamann	Universiteit van Amsterdam
	prof. dr. H.J. Honing	Universiteit van Amsterdam
Overige leden:	prof. dr. E.O. Aboh	Universiteit van Amsterdam
	prof. dr. J.E. Rispens	Universiteit van Amsterdam
	prof. dr. J.J.M. Schoonen	Radboud Universiteit Nijmegen
	dr. D. Sammler	Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig
	dr. M.J. Henry	Max Planck Institute for Empirical Aesthetics, Frankfurt

Faculteit der Geesteswetenschappen

Dit proefschrift is tot stand gekomen binnen een samenwerkingsverband tussen de Universiteit van Amsterdam en Heinrich-Heine-Universität met als doel het behalen van een gezamenlijk doctoraat. Het proefschrift is voorbereid in de Faculteit der Geesteswetenschappen van de Universiteit van Amsterdam en de Faculteit der Geesteswetenschappen van de Heinrich-Heine-Universität.

This thesis was prepared within a partnership between the University of Amsterdam and the Heinrich-Heine-Universität with the purpose of obtaining a joint doctorate degree. The thesis was prepared in the Faculty of Humanities at the University of Amsterdam and in the Faculty of Arts and Humanities at the Heinrich-Heine-Universität.

Table of Contents

Acknowledgements

First and foremost, I want to thank Silke Hamann, without whom none of this work or my thesis would have been possible. I don't even know how to start thanking you. You have been with me since the first time that I set foot into a university and you have been an encouraging supervisor, mentor and friend ever since. You have introduced me to science, research and academia and you have taught me almost everything I know in these fields. You even unintentionally brought me to my research topic when assigning the classic text 'On Alternating Sounds' by Frans Boas, in which he mentions soundblindness. Thank you for believing in me then and continuing to do so. You have been the most supportive supervisor a PhD student could wish for. Thank you so much!

I was so lucky to not only have one wonderful supervisor, but rather a few of them who have all been incredibly helpful and encouraging (in chronological order of joining my supervisor team):

Peter Indefrey, thank you for giving me the freedom that I needed but also asking critical questions. Paul Boersma, thank you for accepting me into your research group and always setting an example on how to be a good scientist. Henkjan Honing, thank you for challenging my ideas and confronting me with different perspectives.

Thank you to all my lovely colleagues that I have shared an office with over the years in Düsseldorf or Amsterdam (in sort of a chronological order) Marion Krause-Burmester, Melani Wratil, Sophie ter Schure, Dinah Baer-Henney, Kim Strütjen, Vicky Tsouni, Jessica Nieder and Mirjam de Jonge - thank you all for always being there, providing good advice, coffee, chocolate, food and listening to my rants - professional and otherwise.

Thank you also to the other members of the BiPhon group for accepting and welcoming me in your middle in 2013 and for introducing me to Dutch culture (besides the people already mentioned above): Jan-Willem van Leussen, Katrin Wanrooi, Katja Chaldkova, David Weenink, Dirk-Jan Vet, Klaas Seinhorst and Jeroen Breteler. A special thanks to Mirjam and Jeroen for all the countless hours of coffee and good conversation, academic and otherwise. Also special thanks to Mirjam, Silke and Paul for making Amsterdam a second home for me for a while and always having a place for me to stay.

Thank you to Dirk-Jan Vet for the invaluable technical assistance. I honestly don't know what I would have done without you! Thank you so much for all your incredible help, good ideas and interesting conversations!

Thank you to Ruben van de Vijver for also making me part of your research group and also always having an open door and for being supportive. Thank you to Frauke Hellwig for teaching me how to program. Thank you to Koen Van Hooste for coming a long part of this way with me and for all the late night paper corrections. Thank you to Kevin Lüthy for always encouraging and challenging me. Thank you to Jessica Nieder for many good conversations and making organizing things together fun. Thank you to Kathi Sternke, who along with Silke, has come all this way with me and has been my friend since my first day at a university.

Thank you to my student assistants that I have had over the years for the tireless data collection. And a very big thank you to all my participants without whom none of this work would have been possible.

Natürlich möchte ich nicht nur meinen wunderbaren KollegInnen und FreundInnen danken, sondern auch meiner Familie. Danke Mama, dass Du mich immer das hast machen lassen, was ich für richtig hielt, und mich dabei unterstützt hast! Danke an meine Oma für ihre Geduld und Liebe. Danke an meinen Onkel, Jürgen Moer, dafür, dass Du mich stundenlang geführt hast (auf dem Pferd) und für deine Dickköpfigkeit. Danke, dass ihr drei immer an mich geglaubt habt und mir beigebracht habt, unabhängig zu sein und immer meine Wege zu gehen und dass ihr mir dies ermöglicht habt. Ihr erlebt zwar nicht mehr, wie ich meine Promotion beende, ihr hattet aber trotzdem einen großen Einfluss auf ihr Gelingen und mein Leben. Ich weiß, dass ihr stolz auf mich wärt.

Danke an Henry und Magret Güth, die mir beide schon immer mit Rat und Tat zur Seite stehen.

Danke an Christine und Thomas Roß für die warme Aufnahme in eure Familie und für die Unterstützung in der Endphase meiner Promotion. Danke an Elfriede und Heinrich Wittig für die Ruhe auf der Trift, die zum Abschließen meiner Dissertation essenziell war. Danke an Gisela Roß, dafür, dass Du meine Ersatzoma geworden bist und an Elvira Matthäi für die Versorgung mit Nervennahrung.

Und natürlich möchte ich mich auch bei Tobi bedanken. Du hast mir in so vielen Lebenslagen geholfen und auf so viele Arten dazu beigetragen, dass ich diese Arbeit in Frieden abschließen konnte. Danke, dass Du immer für mich da bist.

Author Contributions

1. Introduction

Jasmin Pfeifer wrote and revised the introduction with feedback from Silke Hamann.

2. The Diagnosis of Congenital Amusia

Chapter 2 is based on two published articles:

Pfeifer, Jasmin & Silke Hamann (2015). Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia, *Frontiers in Human Neurosciences 9:161*.

Pfeifer, Jasmin & Silke Hamann (2015). Web-based testing of congenital amusia with the Montreal Battery of Evaluation of Amusia, *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, eds. Ginsborg, J., Lamont, A., Phillips, M., Bramley, S.

JP posed the research question, recruited and tested the participants, programmed the laboratory-based experiment and chose and performed the statistical analyses and wrote a first draft of the manuscript. Both JP and SH discussed and interpreted the results together, and together restructured and rewrote the text.
The provided explanation applies to both articles.

3. Vowel Perception in Congenital Amusia

Chapter 3 is yet to be submitted.

Pfeifer, Jasmin & Silke Hamann
(to be submitted)

JP recruited and tested the participants, chose and performed the statistical analyses and wrote a first draft of the manuscript. Both JP and SH posed the research question, designed the experiments together, discussed and interpreted the results together, and together restructured and rewrote the text.

4. Word Stress Perception by Congenital Amusics

Chapter IV is based on an accepted book chapter.

Pfeifer, Jasmin & Silke Hamann. Word Stress Perception by Congenital Amusics
(accepted as a chapter in: *How Language Speaks to Music,* eds: Mathias Scharinger, Richard Wiese; Mouton de Gruyter.)

JP recruited and tested the participants, chose and performed the statistical analyses and wrote a first draft of the manuscript. SH created the stimuli. Both JP and SH posed the research question, designed the experiments together, discussed and interpreted the results together, and together restructured and rewrote the text.

5. The Nature and Nurture of Congenital Amusia: A Twin Case Study

Chapter 5 is published as:
Pfeifer, Jasmin & Silke Hamann (2018). The Nature and Nurture of Congenital Amusia: A Twin Case Study, *Frontiers in Behavioural Neurosciences 12: 120*.

JP posed the research question, recruited and tested the participants, chose and performed the statistical analyses and wrote a first draft of the manuscript. Both JP and SH designed the experiments together, discussed and interpreted the results together, and together restructured and rewrote the text.

6. Conclusion

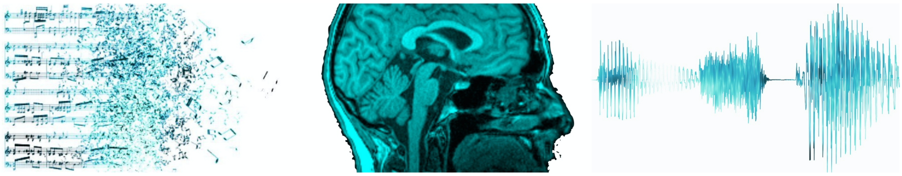
Jasmin Pfeifer wrote and revised the conclusion with feedback from Silke Hamann.

Funding

The research described in this thesis was performed at the Amsterdam Center for Language and Communication (ACLC) at the UvA and the Institute for Language and Information at the Heinrich Heine University, Düsseldorf.

The research for/publication of this doctoral thesis received financial assistance from grant 322-75-004 from the *Netherlands Organisation for Scientific Research* (NWO) to Paul Boersma and Jasmin Pfeifer.

In addition, two workshops pertaining to the topic of the thesis were organized by Silke Hamann and Jasmin Pfeifer. These workshops were financed by SMART grants from the UvA.



1. Introduction

A Minor Bird

I have wished a bird would fly away,
And not sing by my house all day;
Have clapped my hands at him from the door
When it seemed as if I could bear no more.
The fault must partly have been in me.
The bird was not to blame for his key.
And of course there must be something wrong
In wanting to silence any song.

Robert Frost

The topic of this thesis is congenital amusia and how this disorder influences different aspects of affected individuals' perception. In the present chapter, I first introduce the concept of congenital amusia, its symptoms and possible underlying causes and its implications for speech perception. The following chapters all contain their own introductions, discussions and conclusions, hence the relative brevity of this chapter and of the conclusion in Chapter 6.

In the next chapter, the current diagnosis criteria for congenital amusia are reviewed and challenged (Chapter 2). Next, vowel perception in German congenital amusics is investigated with a behavioral and an electrophysiological study (Chapter 3). This is followed by an investigation of word stress processing in German amusic, again using a behavioral and an electrophysiological study (Chapter 4). Finally, a detailed assessment of a dizygotic twin pair, one member of which is amusics, gives insights into the nature versus nurture aspects of congenital amusia (Chapter 5). This dissertation is concluded by a summary of the findings and of the new questions that arose during the research that warrant further investigation (Chapter 6).

Congenital Amusia is an innate disorder that has long been characterized as negatively affecting music perception, hence its name (Peretz, 2001) and its various denominations throughout history: Note deafness (Allen, 1878); sound blindness (LeConte, 1887), tone deafness (Cox, 1948), tune deafness (Kalmus 1948), dysmelodia (Klamus & Fry, 1980) or dysmusia (Geschwind, 1984). This disorder is not caused by hearing loss, brain damage or insufficient music exposure (Ayotte, Peretz & Hyde, 2002) and about 1.5% (Peretz & Vuvan, 2017) to 4% (Kalmus & Fry, 1980) of the general population are said to be affected, depending on the diagnosis criteria. Furthermore, a hereditary pattern is assumed (Peretz, Cummings & Dube, 2007).

Most of the above mentioned, historical texts describe single cases of extremely unmusical individuals and detail different aspects that these people struggled with. Taken together, they yield a surprisingly accurate picture of the musical symptoms of amusics that are described in detail below. Overall, the picture of an individual struggling with music in all its aspects in many different social contexts emerges. Even though the social implications of amusia are not the topic of this thesis, it ought to be noted that individuals with amusia feel great social pressure concerning all music-related activities, and as noted in Frost's poem above, always feel there is "something wrong" with them when it comes to music. Especially Geschwind (1984) points out that the degree to which amusics suffer depends on the demands their society places on them. He points out that dyslexia was not too problematic until the spread of literacy in our society, while amusia, or as he called it, dysmusia, would have been a far greater problem in societies in which music played a greater role in social contexts. Geschwind's (1984) example is the reprieve from poverty through participation in a choir, but more activities, such as participation in communal singing and rituals come to mind. And as an extension of this, communities that make use of emulated speech systems, i.e. systems in which spoken language is transformed into whistles (e.g. Mazateco whistle speech, Cowan, 1948) or musical sounds such as drum beats (e.g. Bora drummed language: Seifart, Meyer, Grawunder & Dentel, 2018), would be highly problematic for amusics, and amusia might lead to their social exclusion.

None of the early accounts relayed any speech perception impairments. It is, however, conceivable that speech impairments might have gone undetected, as they have in the more recent literature in the first few studies. That might partly be due to the native languages of the investigated amusics: French and English. Non-amusic native speakers of French are reported to be stress-deaf due to the absence of contrastive stress in this language (Dupoux, Pallier, Sebastian & Mehler, 1997; Dupoux, Sebastián-Gallés, Navarrete & Peperkamp, 2008) and native English speakers have been reported not to process stress to identify lexical items due to the small number of minimal pairs differentiated only by stress (Cutler, 1986). In languages that utilize contrastive word stress or tone, one would expect reports of individuals who struggle with it.

A very early anecdotal account, stemming from classic Vedic literature, warns against the disastrous consequences of the wrong utterance of word stress. In the Satapatha Brahmana (1.6.3.8; translated by Eggeling, 1882), a text describing Vedic rituals and mythology dating back to around 700 BCE, a verse warns the reader of faulty word stress and the dire consequences it might have: The demon Tvashtri who felt slighted by the god Indra performed a ritual wanting to create a child that would kill Indra by naming it *Indraśatru* 'slayer of Indra' (underscore denotes the stressed syllable). However, he mispronounced the expression as *Indraśatru* 'slayed by Indra', which resulted in his son being killed by Indra.

This is, most likely, not an account of an amusic individual. However, it might point to the fact that people who struggled with the right accentuation of words in languages such as Vedic, in which stress is distinctive, might have existed and might have had trouble or might have been perceived as simple minded. It is therefore surprising that a systematic investigation of this disorder, its musical and speech related symptoms, started as late as 2001 by Isabelle Peretz and colleagues. Peretz (2001) named the disorder congenital amusia, after its acquired counterpart. Amusia most likely received its denomination from the ancient Greek *amousos,* referring to musical anti-heroes “marked by a deficit of culture” (Harmon 2003:351), which is yet another example of the extremely negative depiction of people with musical deficits throughout history.

Congenital amusia, as it is described today, causes lifelong deficits in pitch and partly also rhythm perception, and the most apparent symptoms to the affected individuals themselves are various inabilities in the musical domain such as: An inability to recognize familiar melodies, an inability to detect of out-of-tune notes or off-key singing, and an inability to clap along or to synchronize with a beat. Possibly due to those clear symptoms, early research has mostly focused on its influence on music. Hence, congenital amusia has long been characterized as a music-specific disorder (Ayotte, Peretz & Hyde, 2002; Peretz et al., 2002; Peretz et al., 2002; Peretz, Blood, Penhune & Zarorre, 2001). Different aspects of musical engagement have been assessed and found impaired in amusia, such as pitch perception (Peretz et al., 2002), pitch production (Dalla Bella, Berkowska & Sowiński, 2011), rhythm perception (Foxton, Nandy & Griffiths, 2006), beat synchronization (Sowiński & Dalla Bella, 2013), timbre perception (Marin, Gingras & Stewart, 2012), consonance rating (Ayotte et al., 2002), and musical emotion perception (Marin, Thompson & Stewart, 2012). The underlying cause of this multi-faceted disorder has been hypothesized to be a fine-grained pitch processing deficit (Ayotte et al., 2002), a pitch memory deficit (Gosselin, Jolicœur & Peretz, 2009), a statistical learning deficit (Peretz, Saffran, Schön & Gosselin, 2012) or a rapid-auditory processing deficit (Williamson, McDonald, Deutsch, Griffiths & Stewart, 2010). There is no consensus yet, and a multi-causal deficit is most likely responsible for the different symptoms exhibited by amusics. The pitch perception deficit, which is clearly present, has been hypothesized to be due to either a processing or a memory deficit.

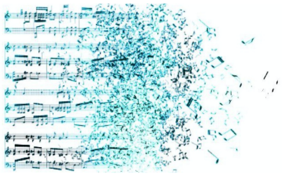
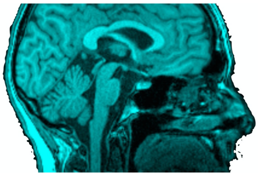
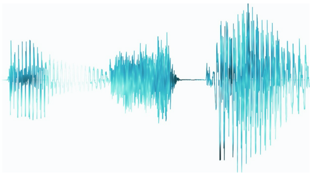
Regardless of the underlying deficit, pitch perception also plays an important role in language perception, and recently more attention has been paid to possible pitch perception impairments in speech due to amusia (Patel, Foxton & Griffiths, 2005; Patel, Wong, Foxton, Lochy, & Peretz, 2008; Jiang, Hamm, Lim, Kirk & Yang, 2010; Liu, Patel, Fourcin & Stewart, 2010; Hamann, Exter, Pfeifer & Krause-Burmester, 2012). Pitch is important in the transfer of linguistic meaning. In intonation it is, for example, used to disambiguate questions from statements or for emphasis; on the word level it is used to distinguish words with similar segmental structure but different stress patterns (e.g. English *present* vs. *present*, where underscore denotes the stressed syllable) or to distinguish words with identical segments but different tones (e.g. in tone languages such as Mandarin Chinese). Such word stress differences served as the basis for the investigation in Chapter 4.

Other areas of speech perception have also been shown to be affected by amusia, e.g. emotional prosody in language (Lolli, Lewenstein, Basurto, Winnik & Loui, 2015; Thompson, Marin & Stewart, 2012) or tone language perception (Liu, Jiang, Wang, Xu & Patel, 2015; Liu, Maggu, Lau & Wong, 2015; Liu, Xu, Patel, Francart & Jiang, 2012b; Tillmann et al., 2011a). Taken together, all of these findings have evoked a change in how this disorder is seen: Previously it was described as domain-specific to music, whereas now it is viewed as a domain-general disorder affecting pitch processing (Hamann et al., 2012; Liu et al., 2010; Zhang, Shao & Huang, 2017). So far, the study of speech perception impairments caused by amusia has almost exclusively focused on areas involving pitch as a perceptual correlate of speech sounds, which is probably due to the fact that most hypotheses about the underlying deficit of amusia are based on some form of pitch perception deficit. Speech, however, also makes extensive use of other information in the speech signal such as spectral frequencies. The latter are especially relevant in the perception of vowels. First reports on vowel perception in congenital amusia have recently appeared (Huang, Zhang, Shi, Yan, & Wang, 2016; Zhang et al., 2017; Tang et al., 2018). In all of these studies, vowel perception has only been investigated in conjunction with tone in tonal languages such as Mandarin or Cantonese. The perception of vowels in a non-tone language, namely German, is the topic of Chapter 3.

Before conducting our speech perception studies, we realized that it was necessary to re-evaluate the current practice of diagnosing amusics. In order to properly design and conduct studies about congenital amusia, one needs a subject pool diagnosed with a reliable tool, clear cut-off scores and adequate scoring methods. Currently, the only available tested tool is the *Montreal Battery of Evaluation of Amusia* (MBEA; Peretz, Champod & Hyde, 2003). Its usage and usefulness are therefore reviewed in the following chapter (Chapter 2). Different aspects such as cut-off scores, number of used subtests, manner of testing, employed statistics and the resulting prevalence and subtypes are discussed. Suggestions are made for a more appropriate usage, which are then carried out in the following chapters.

Lastly, in Chapter 5, a twin case study is presented. One twin was diagnosed as amusic, while the other is not amusic. The twin pair is thoroughly assessed with a battery of tests ranging from musical and speech perception tasks to spatial tasks. The findings give rise to a debate about nature versus nurture aspects of this disorder and opens up questions concerning the heritability of the disorder.

All in all, the goal of this thesis was to untangle certain aspects of auditory perception in congenital amusia by investigating its cognitive, neural and genetic underpinnings. While some questions – the ones that led to the research design of each of the following chapters – have received answers, many new questions arose. These findings as well as the newly formed question are summarized in the last chapter (Chapter 6) of this thesis.

2. The Diagnosis of Congenital Amusia

<u>This chapter is based on:</u>

Pfeifer, Jasmin & Silke Hamann (2015). Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia, *Frontiers in Human Neurosciences 9:161*.

Pfeifer, Jasmin & Silke Hamann (2015). Web-based testing of congenital amusia with the Montreal Battery of Evaluation of Amusia, *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, eds. Ginsborg, J., Lamont, A., Phillips, M., Bramley, S.

Abstract

This article presents a critical survey of the prevalent usage of the *Montreal Battery of Evaluation of Amusia* (MBEA; Peretz et al., 2003) to assess congenital amusia, a neuro-developmental disorder that has been claimed to be present in 4% of the population (Kalmus and Fry, 1980). It reviews and discusses the current usage of the MBEA in relation to cut-off scores, number of used subtests, manner of testing, and employed statistics, as these vary in the literature. Furthermore, data are presented from a large-scale experiment with 228 German undergraduate students who were assessed with the MBEA and a comprehensive questionnaire. This experiment tested the difference between scores that were obtained in a web-based study (at participants' homes) and those obtained under laboratory conditions with a computerized version of the MBEA. In addition to traditional statistical procedures, the data were evaluated using *Signal Detection Theory* (SDT; Green and Swets, 1966), taking into consideration the individual's ability to discriminate and their response bias. Results show that using SDT for scoring instead of proportion correct offers a bias-free and normally distributed measure of discrimination ability. It is also demonstrated that a diagnosis based on an average score leads to cases of misdiagnosis. The prevalence of congenital amusia is shown to depend highly on the statistical criterion that is applied as cut-off score and on the number of subtests that is considered for the diagnosis. In addition, three different subtypes of amusics were found in our sample. Lastly, significant differences between the web-based and the laboratory group were found, giving rise to questions about the validity of web-based experimentation.

1. Introduction

Congenital amusia is a perceptual disorder that affects music and speech perception. Congenital amusics do not suffer from a hearing deficit nor do they have any form of brain lesion (Ayotte, Peretz and Hyde, 2002). Rather, the disorder is an innate one and the exact neural underpinnings are still under investigation. Therefore, no neurological markers can be used to diagnose amusia. Instead, research has revealed several behavioral markers, such as pitch perception deficits and a pitch memory deficit. The main tool used to diagnose amusia nowadays is the *Montreal Battery of Evaluation of Amusia*

(MBEA; Peretz, Chambod and Hyde, 2003), which was originally developed to confirm acquired amusia in patients with brain lesions.

In the present study, we first describe the set-up of the MBEA (section 1.1.) and give an overview of its current usage and limitations (section 1.2.). Section 2. presents a large-scale study that compares web-based with laboratory-based usage of the MBEA, and evaluates the MBEA scores with data on musical performance additionally obtained with a questionnaire. The results of this experiment are presented in section 3. A discussion of the results is given in section 4.

1.1. The Montreal Battery of Evaluation of Amusia

The MBEA is a test battery developed with the main objective of assessing the musical abilities of brain-damaged patients that suffer from acquired amusia, but is nowadays used to diagnose congenital amusia. It consists of six subtests, three of which test melodic organization (scale, contour and interval subtest), two test temporal organization (rhythm and meter subtest) and one tests melodic memory (memory subtest), based on a model of music processing summarized by Peretz and Coltheart (2003).

All six subtests use a selection of musical phrases that were specifically composed for this purpose according to the principles of the Western tonal system. These phrases are monophonic, i.e. they consist of a single voice, and they last 3.8 to 6.4 seconds (mean of 5.1 s) for all but the metric test, where they are polyphonic and twice as long (with a mean of 11 s). The procedure is the same for the first four subtests (scale, contour, interval and rhythm): The participants are presented with two practice trials and 31 experimental trials. A trial consists of a target melody and a comparison melody (thus a stimulus pair), which are separated by a 2-second silent interval. Each trial is preceded by a warning tone and followed by a 5-second silent interval. 15 trials have comparison melodies that are identical to the target melody and 15 have comparison melodies that are altered in one note with respect to the target melody: In the scale subtest, the altered melodies violate the key but keep the overall contour intact; in the contour subtest, they violate the contour while keeping the key intact; and in the interval subtest, key and contour are kept intact but the pitch interval is violated. For the rhythm subtest, the rhythmic grouping of the comparison melody is changed by altering the duration of two adjacent notes. In addition to those 30 trials, each subtest contains a catch trial to ensure that the participants are paying attention and not simply guessing. For the catch trial, the pitch of the comparison melody was changed randomly, so that there is a clearly noticeable difference between the two melodies. For the first four subtests, participants are asked whether the two melodies they hear are the same or different, following an AX design. The last two subtests (meter and memory) follow a different design. In the meter subtest, 30 two-phrase sequences in duple or triple meter are used, and the participants have to judge whether the presented melody is a march or a waltz. The memory subtest presents again only single melodies, half of which already occurred in the previous subtests, the other half is new, and participants have to indicate for each melody whether they have heard it before during the previous subtests or whether it is new.

The MBEA was used by Peretz et al. to test 160 participants without known neurological problems, who were not selected for musical ability. For each participant, the number of correct responses per subtest and an average score of the six subtests was calculated. As cut-off scores for congenital amusia, Peretz et al. propose 2 standard deviation (SD) below the mean of the 160 participants, thus an average score below 21.6, or 76.6%, cf. Table 2.1.

	Scale	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**	**Average**
Mean correct responses	27	27	26	27	26	27	27
Standard deviation (SD)	2.3	2.2	2.4	2.1	2.9	2.3	1.6
Cut-off score (mean – 2 SD)	22	22	21	23	20	22	21.6*
Cut-off in %	73.3	73.3	70	76.7	66.7	73.3	72.2*
% with perfect score	17	9	7	15	14	10	3
% below cut off	3	1	1	1	1	1	2

Table 2.1: MBEA test scores for the six subtests and average score in the study by Peretz et al. (2003: 66).
*Peretz et al. (2003: 69) list an average cut-off score of 23 and a cut-off percentage of 78%, which is probably due to a rounding error (see Wise 2009: 115).

According to Peretz et al. (p. 65), the MBEA subtests provide a *sensitive* measure since less than 20 percent of the participants obtain perfect scores for each subtest, and only 3% of the participants obtained a perfect score for all subtests (see Table 2.1 row 5), while less than 2% (3 participants) had average scores that were below 2 SD of the mean (Table 2.1 row 6). These average scores approximate a normal distribution, though the scores for the individual subtests display a skew to the right. Peretz et al. furthermore state that the MBEA displays *test-retest reliability*, based on a retest of 28 participants four months after initial testing, though the performance of these participants improved (p. 66).

(Peretz et al., 2003) *validated* the MBEA with two subtests (melody and meter) of the *Musical Aptitude Profile* (MAP; Gordon, 1965), a test battery widely used in North America to test musical abilities. These two subtests, which were chosen because they were closest in content and format to the MBEA, were administered to 68 subjects. These participants obtained similar levels of performance for the MBEA and the MAP, and the two scores positively correlated ($r = 0.53$, $p < 0.01$).

1.2. Applications and limitations of the MBEA

Currently most studies investigating congenital amusia utilize the MBEA, including those performed by researchers who are not associated with Peretz' research group (Foxton et al., 2004; Patel et al., 2005; Foxton et al., 2006; Douglas and Bilkey, 2007; Mandell et al., 2007; McDonald and Stewart, 2007; Loui et al., 2008; Patel et al., 2008; Loui et al., 2009; Nguyen et al., 2009; Tillmann et al., 2009; Wise, 2009; Jiang et al., 2010; Liu et al., 2010; Williamson et al., 2010; Williamson and Stewart, 2010; Loui et al., 2011; Omigie and Stewart, 2011; Tillmann et al., 2011a; Tillmann et al., 2011b; Williamson et al., 2011; Hamann et al., 2012; Jiang et al., 2012a; Jiang et al., 2012b; Liu et al., 2012a; Liu et al., 2012b; Loui and Schlaug, 2012; Omigie et al., 2012a; Omigie et al., 2012b; Thompson et al., 2012; Tillmann et al., 2012; Williamson et al., 2012; Albouy et al., 2013a; Albouy et al., 2013b; Jiang et al., 2013; Liu et al., 2013; Omigie et al., 2013; Launay et al., 2014; Pfeifer and Hamann, 2014).

The actual application of the MBEA differs in terms of number of subtests, items, and cut-off scores that are employed, in their mode of testing (web-based or in the laboratory) with or without additional questionnaire, in their predictions on the prevalence of amusia, and in whether they differentiate subtypes of amusia. In the following subsections, we summarize and discuss the different usages found in the literature.

1.2.1. Scoring and subtests

All studies testing congenital amusia calculate a score based on the sum of correct answers without distinguishing between different types of stimuli or answer categories. They usually also calculate an average score but include different numbers of subtests.

Isabelle Peretz and her colleagues use all six subtests of the MBEA (Hyde and Peretz, 2003; Hyde and Peretz, 2004; Peretz et al., 2005; Hyde et al., 2006; Hyde et al., 2007; Peretz et al., 2007; Peretz et al., 2008; Moreau et al., 2009; Peretz et al., 2009; Hutchins et al., 2010a; Hutchins et al., 2010b; Nan et al., 2010; Hyde et al., 2011; Cousineau et al., 2012; Mignault Goulet et al., 2012; Peretz et al., 2012; Hutchins and Peretz, 2013; Moreau et al., 2013; Phillips-Silver et al., 2013). These studies use the scores by Peretz et al. (2003) as cut-off score. In the early studies by Ayotte et al. (2002) and Peretz et al. (2002) a cut-off score of 3 SD below mean was used. As already pointed out by Wise (2009: 43), it is not clear why this change from 3 to 2 SD was made, but it resulted in more people being assessed as having amusia.

The research group led by Lauren Stewart and her colleagues uses only the first four subtests of the MBEA and calculates the sum of the first three, pitch-based, subtests (McDonald and Stewart, 2007; Liu et al., 2010; Williamson et al., 2010; Williamson and Stewart, 2010; Omigie and Stewart, 2011; Williamson et al., 2011; Liu et al., 2012a; Omigie et al., 2012a; Omigie et al., 2012b; Thompson et al., 2012; Williamson et al., 2012; Liu et al., 2013; Omigie et al., 2013). As cut-off score, they use 65 out of 90 correct answers on the first three subtests (72% correct).

Several large-scale studies using all six subtests of the MBEA employ cut-off scores that are based on the means they obtain for their own participants. Cuddy et al. (2005)'s 100 control participants (subjects who reported not to be tone deaf) achieved lower mean correct responses than the control group by Peretz et al. (87% compared to 91%). As a result, Cuddy et al. set their cut-off scores at 2 SD below the mean of their controls, thus at 72%, resulting in 3–5% of the participants being diagnosed as amusic (as opposed to 18% with Peretz et al.'s cut-off scores). These scores are much lower than the ones obtained by Peretz et al., with the exception of the score for the memory test, see Table 2.2 rows 1 and 2 compared to the last two rows. Wise (2009), who uses 24 test items per subtest instead of 30, also employs cut-off scores that lie 2 SD below the means of her own 24 controls (participants without self-reported problems in music perception and performance). These scores were mostly lower than the ones used by Peretz et al., cf. Table 2.2 rows 3 and 4. In a study on the presence of amusia in native speakers of a tone language, Nan, Sun and Peretz (2010) tested 117 Mandarin speakers with no self-declared musical problems. Their cut-off scores are also given in Table 2.2 (rows 5 and 6).

Source		**Scale**	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**	**Average**
Cuddy et al. N = 100	Cut-off scores	20.1	19.4	18.6	20.2	15.1	22.7	21.5
	Cut-off in %	67.0	64.7	62.0	67.3	50.3	75.7	71.7
Wise N = 24	Cut-off scores	21.5	19.7	19.8	23.2	19.4	20.8	22.4
	Cut-off in %	71.5	65.7	66.1	77.4	64.6	69.3	74.6
Nan et al. N = 117	Cut-off scores	19.3	20.9	17.7	22.0	16.2	21.5	21.5
	Cut-off in %	64.2	69.6	59.0	73.3	53.9	71.8	71.7
Peretz et al. N = 160	Cut-off scores	22	22	21	23	20	22	21.6
	Cut-off in %	73.3	73.3	70	76.7	66.7	73.3	72.2

Table 2.2: MBEA cut-off scores for the six subtests and the average score by control subjects in the large-scale studies by Cuddy et al. (2005), Wise (2009), Nan et al. (2010), and Peretz et al. (2003) (for comparison). Not all studies provided cut-off scores both in absolute numbers and in percentage, the missing data were calculated by the present authors.

These percentages are comparable to the ones by Wise's control participants (though markedly lower for the meter subtest) and thus also lower than the original scores proposed by Peretz et al.

Almost all studies use *average cut-off scores* to diagnose amusics, i.e. the performance on an individual subtest does not matter as much, especially when 6 subtests are used. An example for this is the study by Ayotte et al. (2002), where the average score of every amusic is 3 SD below the mean of the controls, but when considered on individual subtests, none of the 11 amusics failed all the subtests and some scored below the cut-off for only 2 subtests.

The practice of adding up all correct responses to calculate a score for the MBEA is criticized by Henry and McAuley (2013), as it might misdiagnose people as amusics who have a large response bias but normal discriminatory abilities. They propose the use of *Signal Detection Theory* (SDT) (Green and Swets, 1966; Macmillan and Creelman, 2005), which is a psychophysical approach to measuring performance that takes into account the individual's response bias and their ability to discriminate, both important considerations for testing a population with a perceptual deficit. Henry and McAuley (2013) compare the performance of participants who completed the standard MBEA with the performance of participants who additionally had to rate how confident they were of their answers. With these confidence scores, Henry and McAuley computed SDT scores and found a potential misclassification of 33%.

A possible *misdiagnosis* of amusics could be ascribed to a high rate in Type II error, thereby including individuals with a large response bias who have otherwise normal perceptual abilities. This would mean that by using SDT, a more rigorous standard of diagnosis would be employed, leading to fewer Type II errors. This consideration is especially important when (re-)assessing studies that have obtained null results. It seems possible that these studies included a large group of misdiagnosed individuals, thereby tainting the results.

Wise (2009) and Henry and McAuley (2010) point out the *negative skew* in the distribution of scores on the individual subtests, and furthermore that most studies using the MBEA apply parametric statistics without testing whether their data are normally distributed (exceptions are Douglas and Bilkey, 2007; Provost, 2011).

Some studies use MBEA subtests for *screening*, which could lead to potentially higher MBEA scores in the later testing; recall the improved performance by participants who were retested after four months in the study by Peretz et al. (2003: 66). Such a potential learning effect for participants screened with MBEA subtests hinders the interpretation and cross-study comparison of reported final scores.

As we could see, there is no agreement on the cut-off scores and the number of employed subtests. Both vary considerably across studies, which makes cross-study comparisons difficult if not impossible. In order to employ the MBEA as diagnostic tool, a standardized usage would be necessary.

1.2.2. Web-based versus laboratory testing

In recent years, web-based research has become more and more common. While the MBEA is mostly conducted in a laboratory, some studies on congenital amusia employ web-based MBEA subtests for pretesting, e.g. Lauren Stewart and colleagues, who use a web-based pretest consisting of the scale and the rhythm subtest of the MBEA.

Peretz et al. (2008) proposed a *web-based amusia test* based on the MBEA (see e.g. Liu et al., 2010, 2013; Williamson & Stewart, 2010). This test consists of 3 conditions with a total of 72 melodies based on 12 melodies from the MBEA. The task of the experiment was to spot incongruities that were inserted in these melodies and not a comparison of two melodies, as in the MBEA. In one condition, off-beat tones or silences were inserted, thereby altering the meter of the phrase. In the other two conditions, a mistuned note or an out-of-key note were inserted, respectively. Peretz and colleagues used the MBEA, on which this test is based, to validate it by correlating the scores on the MBEA subtests with scores in these subtests. Similar to the MBEA, the off-beat test is shown not to be normally distributed. The average score is described, but not statistically shown, to be normally distributed, while visual inspection of the provided material also reveals a skew in the data. They also mention discrepancies between the two tests: 19% of people diagnosed as amusic would not be diagnosed as such with the online test. This result contradicts the expectations that participants tested online should perform equally well or slightly worse than lab-tested participants due to uncontrolled testing conditions (such as noise, unrestricted amount and length of breaks, etc.). Peretz et al. explain their findings with the difference in task between the two tests: For the MBEA (tested in the lab), participants have to compare melodies, which is more demanding than the online test of detecting incongruities, because it requires participants to hold pitch information in their working memory, while the web test does not involve working memory.

A discrepancy between web-based results and laboratory results has often been observed in psychological research. Krantz and Dalal (2000) comment that this does not demonstrate a lack of validity of web-based experiments, since most variables seem not to be influenced by varying environments. However, they also point out that auditory research is an exception to this observation as a stable and quiet environment is crucial for the success of this type of experiment. For the assessment of amusia with a web-based version of the MBEA, this could mean severe misdiagnoses of participants.

1.2.3. Use of additional questionnaires

In addition to testing with the MBEA, many studies report the usage of a *questionnaire* pertaining to information on general education, music education, language background and musical performance such as singing and dancing (Cuddy et al., 2005; Wise, 2009; Provost, 2011). In most cases where a questionnaire was used, it is not reported how it is analyzed in relation to the MBEA results (e.g. Ayotte et al., 2002; Peretz et al., 2007; Liu et al., 2012b). One of the exceptions is Peretz et al. (2008) with 101 items on demographic and music-related information. However, only correlations between a small number of questionnaire items (age, gender, years of education and music training) and MBEA test scores are reported.
Questionnaires could in principle provide valuable additional information in the assessment of amusics with the MBEA, but in order to evaluate their contribution more studies are required that systematically analyze the correlation between the questions used and the MBEA scores.

1.2.4. Prevalence

The MBEA is also used to estimate the *prevalence of congenital amusia* in the general population. Most amusia studies state a prevalence of 4%, referring to Kalmus and Fry (1980) (e.g. Ayotte et al., 2002; Foxton et al., 2004; Cuddy et al., 2005; Mandell et al., 2007; Peretz et al., 2007; Peretz et al., 2009; Liu et al., 2010; Tillmann et al., 2010; Williamson and Stewart, 2010; Omigie and Stewart, 2011; Liu et al., 2012b; Williamson et al., 2012; Omigie et al., 2013). Kalmus and Fry (1980) introduced the *Distorted Tunes Test* (DTT), consisting of 26 well-known tunes to assess congenital amusia (or tone deafness as they called it). Incorrect notes were inserted into 17 of these tunes. Kalmus and Fry's criterion for the presence of amusia was the inability to detect wrong notes in at least three out of the 17 incorrect tunes. They tested 604 adults and based on this data they estimated a prevalence of 4.2%.

Recently, Peretz (2013) stated 2.5% as the prevalence of amusia in the general population, and added that the use of only the MBEA scale subtest by Provost (2011) resulted in a prevalence of 3.2%. Provost (2011) used the online study based on Peretz et al. (2008) described in section 1.2.2. For the 1100 participants who completed the test and fitted the age and education criteria, scores were considered individually and any participant falling below the cut-off score on one of the three subtests was considered amusic. This yielded a total of 11.6% amusics, supporting the observation above that on-line testing yields a higher prevalence of amusia.

Henry and McAuley (2010: 414) point out that the MBEA, just like other methods to assess the prevalence of disorders (including dyslexia and dyscalculia), suffers from an *arbitrary cut-off* problem and that a cut-off of 2 SD from the mean in normally distributed values (as claimed for the average score of the MBEA) would by definition result in a 2.28% expected occurrence rate. The same criticism can be applied to the prevalence proposed by Kalmus and Fry (1980), as their DTT shows the same arbitrary cut-off score and lack of well-established psychometric properties (Ayotte et al., 2002; Hyde and Peretz, 2004; Henry and McAuley, 2010). Henry and McAuley therefore propose to include structured interviews with participants for predictions on prevalence (see section 1.2.3 above).

1.2.5. Subtypes of amusia

Wise (2009) and Henry and McAuley (2010) criticize the widespread use of average scores for the MBEA, because this practice ignores heterogeneous behavior of participants across the six subtests. In the study by Ayotte et al. (2002) for instance, only the scale subtest was failed by all congenital amusics (Wise, 2009: 43). Wise further reports that for the rhythm and the meter subtest, half of the participants usually pass and more than half pass the memory subtest. At the same time, participants have been reported who only have problems with the rhythm subtest (Peretz et al., 2003). All this points to the existence of several subtypes of congenital amusia with a possible dissociation between pitch- and rhythm-related deficits, as already suggested by Peretz et al. (p. 70). Some studies using the MBEA introduce the amusic subtype of *beat deafness* (Phillips-Silver et al., 2011) or *dysrhythmia* (Launay et al., 2014). Phillips-Silver et al. report a single case of rhythmic deficits with intact pitch perception, while Launay et al. identify three such cases. The opposite, intact rhythmic perception with impaired pitch perception, has also been reported by Phillips-Silver et al. (2013).

Reports of a subtype with rhythm deficits are less frequent, possibly due to a low proportion of rhythm-related subtests in the MBEA. Provost (2011) proposes four different subtypes: Pitch-deaf amusics, pitch-perception amusics, pitch-memory amusics and beat-deaf amusics. The latter classification does not include an amusia type that has both a pitch perception and a rhythm perception deficit and rather focuses on different pitch abilities.

This overview shows that even though there is large overlap in the proposed subtypes of amusia, clear-cut definitions for such subtypes are still missing. Furthermore, we can conclude that it is advisable to use cut-off scores of single subtests instead of average scores in order to advance further research on subtypes of amusia.

2. Materials and Methods

2.1. Participants

280 first year undergraduate students in general linguistics at the Heinrich Heine University Düsseldorf participated in our study. The participants were not preselected for the presence or absence of musical disorders such as amusia, or specific levels of musical experience. All participants gave informed written consent to participate in this study and received course credit for their participation. All data were collected in accordance with the declaration of Helsinki. The participants took a hearing test and answered a detailed questionnaire about their linguistic and musical background (experience with and attitude to music and dance, in performance and perception). An intelligence test was not performed as all participants were university students and expected to have an average to high level of intelligence.

A total of 52 students was excluded from data analysis. 8 did not have normal hearing as assessed by pure tone audiometry at 250–8000 Hz. Normal hearing was defined as a mean hearing level of 20 dB or less in both ears. 45 participants had a different native language than German. In order to keep the variance between participants as little as possible, these participants were excluded as well. Of the remaining 228 participants, 117 completed a web-based version of the MBEA at home and 111 a computer-implemented version in a sound-attenuated booth in our laboratory. Participant details can be found in Table 2.3. The last row in this Table shows that the two participant groups did not differ significantly in their characteristics.

Group		**Age**	**Years of education**	**Years of music education**	**Handedness**	**Gender**
Labor-atory	Mean	22.7	14.4	5.9	101 right 7 left 3 ambid.	90 f 21 m
	Range	20 - 35	12-23	0 - 12		
Web-based	Mean	22.0	14.7	6.3	107 right 9 left 1 ambid.	99 f 18 m
	Range	19 - 36	12-22	0-17		
Total	Mean	22.3	14.6	6.1	207 right 16 left 4 ambid.	188 f 39 m
t-test means	*t*	1.82	-1.06	-0.82	-	-
	p	0.71	0.29	0.41	-	-

Table 2.3: Descriptive statistics and results of *t*-tests comparing laboratory and web-based participant characteristics. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

2.2. Procedure

All participants completed the MBEA. Half of them completed a computer-implemented version in a sound-attenuated booth, where the stimuli were presented over AKG K 601 headphones using Praat (Boersma and Weenink, 2011) on a Windows XP computer. These participants could adjust the volume to a comfortable level and had unlimited time. The other half completed a web-based version of the MBEA. This group was informed before the experiment that they should use headphones and take the test in a quiet environment without any distractions. The on-screen instructions for both groups were identical. For the first four and the sixth subtest, participants received two examples with feedback before the beginning of each subtest. For the fifth (meter) subtest, participants received four examples, instructing them what a march and a waltz sound like. A detailed description of the MBEA stimuli and the general procedure was given in section 1.1.

The laboratory group took part in the MBEA, filled in the questionnaire and took the hearing test in the same session, which lasted about 70 minutes. The web-based group completed the MBEA online at home. At a later point, these participants came to the laboratory for the hearing test and to answer the questionnaire. A test administrator was always present to answer clarification questions about the questionnaire. At the end of the session, participants were allowed to ask questions about the nature of the study and a couple of weeks later they were informed about the group results.

2.3. Scoring

The MBEA uses a same-different paradigm for the first four subtests. In such a test design, participants respond to two different types of trials (stimulus pairs) in two different ways: A stimulus pair where the comparison melody differs from the target melody is considered a hit when it is correctly identified, and a miss when it is not correctly identified. A stimulus pair with two melodies that are the same is considered a correct rejection when it is identified as identical, and scored a false alarm when it is incorrectly identified as different, see the overview in Table 2.4.

	Response	
Stimulus pair	Different	Same
Different	Hit (H)	Miss (M)
Same	False Alarm (FA)	Correct Rejection (CR)

Table 2.4: Overview of stimulus types and possible responses

Following Henry and McAuley (2013), these scores were also applied to the metric subtest, where trials with a march were treated as different and trials with a waltz as same stimulus pairs, and to the memory subtest, where trials with already used melodies were treated as different and those with new melodies as same stimulus pairs.

Poor performance on the MBEA can occur for different reasons: It can be caused by a high number of false alarms, a high number of misses or a combination of both. We therefore performed not only a conventional analysis of the MBEA by calculating the sum of correct responses, but also employed the SDT measures *d'*, as a measure of sensitivity, and *criterion location* (*c*), as a measure of participants' response bias. Both rely on hit rate (HR) and false alarm rate (FAR). *d'* is equally dependent on H and FA and allows for the fact that sensitivity should increase when H increases and decrease when FA increases. It is calculated by subtracting the inverse of the normal distribution functions of FA from H, converting them into a standard deviation unit (z-scores), cf. (1), and thereby making the measure comparable across tasks (Macmillan and Creelman 2005). A *d'* score of 0 means a participant is unable to discriminate between stimuli, and the higher the *d'* score (and thus the sensitivity), the better the participant discriminates between stimuli.

(1) $d' = z(HR) - z(FAR)$

(2) *Criterion Location*: $c = -0.5 \cdot (z(HR) + z(FAR))$

The second measure, *c*, is the participants' response bias, i.e. the tendency to favor one of the two possible responses (Macmillan and Creelman, 2005) and is calculated as in (2). Positive *c* values correspond to a tendency to respond 'same' and negative values correspond to a tendency to respond 'different'.

3. Results

Several analyses were performed on the data. First, the results of the web-based group and the group tested in the laboratory were analyzed and compared, see section 3.1. In section 3.2., the cut-off scores of our sample and the prevalence that we found are compared to the cut-off scores by Peretz et al. (2003). Section 3.3. discusses the use of combined subtests for our data. Section 3.4. shows the results when signal detection theory was applied and discusses the differences in prevalences. Resulting subgroups of amusia are discussed in 3.5. This is followed by an analysis of the questionnaire items in relation to the MBEA scores in section 3.6.

3.1. Web-based versus laboratory testing

The web-based group and the group tested in the laboratory were analyzed separately, by computing sum of correct responses, cf. Table 2.5. The mean of correct responses for the web-based group is generally lower than for the lab-tested group (though it is almost identical for the scale subtest) and the web-based group shows more variation, i.e. SD is larger for every subtest.

Group		Scale	Contour	Interval	Rhythm	Meter	Memory	Average
Web-based	Mean	24.97	23.86	23.21	24.87	24.09	26.34	24.56
	SD	3.03	3.38	3.89	3.80	5.29	3.09	3.94
Laboratory	Mean	24.95	24.68	24.32	25.84	26.07	27.51	25.56
	SD	2.73	3.01	3.29	2.64	3.65	1.77	3.09

Table 2.5: Sum of correct responses for the web-based and laboratory groups (absolute numbers, with maximum of 30 per subtest)

Visual inspection of histograms indicated that the data for the individual subtests and for the average of all subtests are not normally distributed, for an example illustration see Figure 2.1.

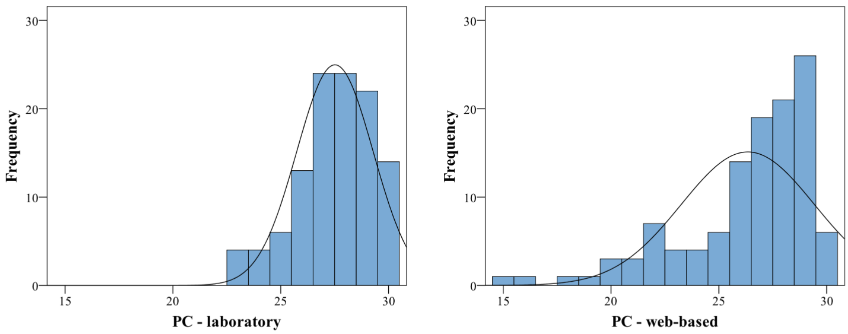


Figure 2.1: Histograms of proportion correct for the Memory subtest for laboratory (left) and web-based group (right). Both exhibit a significant negative skew. For statistics, cf. Table 2.6.

The calculation of skew, kurtosis and Kolmogorov-Smirnov tests yielded significant results as well (for exact values see Table 2.6). All subtest scores and the average scores exhibit a negative skew and are visibly shifted towards the right, indicating a build-up of high scores. Especially the memory subtest in the web-based group exhibits a significant ($p < 0.001$) kurtosis value, indicating that it is not normally distributed. In addition, the Kolmogorov-Smirnov tests yielded significant results as well. The variances between the web-based and the laboratory groups are significantly different for four of the six subtests as revealed by Levene's test and therefore the assumption of homogeneity of variance is also violated for these four tests (for exact values see Table 2.6). For these reasons, additional non-parametric tests were performed. Mann-Whitney-U tests revealed significant differences between the web-based and the laboratory group in four out of six subtests. The contour and interval subtest and the average of all subtests reached significance at $p < 0.05$ and the meter and memory subtests reached significance at $p < 0.01$ (for values see Table 2.6).

Subtest	Group	Skew	SE Skew	z Skew	Kurtosis	SE Kurtosis	z Kurtosis	Kolmogorov-Smirnov Test		Levene's Test		Mann-Whitney-Test (1-tailed)		
								D	*p*	F (1,226)	*p*	U	*z*	*p*
Scale	lab	-1.06	0.23	***-4.62***	1.99	0.46	***4.37***	0.14	***0.00***	1.46	0.23	6239	-0.52	0.30
	web	-0.93	0.22	***-4.18***	0.61	0.44	1.36	0.16	***0.00***					
Contour	lab	-0.70	0.23	***-3.02***	0.05	0.46	0.12	0.12	***0.00***	2.56	0.11	5588	-1.83	*0.03*
	web	-0.45	0.22	*-1.99*	-0.43	0.44	-0.96	0.12	***0.00***					
Interval	lab	-0.81	0.23	***-3.54***	0.55	0.46	1.20	0.14	***0.00***	4.10	*0.04*	5398	-2.21	*0.01*
	web	-0.47	0.22	*-2.09*	-0.40	0.44	-0.90	0.09	*0.01*					
Rhythm	lab	-0.63	0.23	*-2.74*	-0.06	0.46	-0.14	0.14	***0.00***	8.86	*0.003*	5786	-1.43	0.08
	web	-0.96	0.22	***-4.31***	0.28	0.44	0.64	0.15	***0.00***					
Meter	lab	-1.20	0.23	***-5.23***	1.22	0.46	*2.69*	0.17	***0.00***	19.76	***0.00***	5251	-2.51	*0.01*
	web	-0.85	0.22	***-3.80***	0.13	0.44	0.29	0.15	***0.00***					
Memory	lab	-0.66	0.23	*-2.86*	0.05	0.46	0.11	0.15	***0.00***	19.85	***0.00***	5309	-2.41	*0.01*
	web	-1.39	0.22	***-6.22***	1.78	0.44	***4.01***	0.20	***0.00***					
Average	lab	-0.69	0.23	***-3.02***	0.55	0.46	1.20	0.33	***0.00***	8.60	*0.004*	5353	-2.29	*0.01*
	web	-0.94	0.22	***-4.18***	0.62	0.44	1.39	0.28	***0.00***					

Table 2.6: The results of variance and normality analyses, comparison between laboratory and web-based group per MBEA subtest. Bold indicates $p < .001$, italics $p < .05$

3.2. Prevalence

The means of the sum of correct responses, standard deviations and different cut-off scores are given in Table 2.7 for group tested in the laboratory and in Table 2.8 for the web-based group. The *average* values are calculated by averaging the scores of all participants across all subtests, it is not an average of the means or standard deviations. The *pitch average* is an average of the scores for the scale, contour and interval subtest.

Sum of correct responses	Scale	Contour	Interval	Rhythm	Meter	Memory	Average	Pitch Average
Mean	24.95	24.68	24.32	25.84	26.07	27.51	25.48	24.64
SD	2.73	3.01	3.29	2.64	3.65	1.77	2.06	2.47
Cut-off (2 SD)	19.49	18.66	17.74	20.56	18.77	23.97	21.36	19.7
Cut-off %	65.0	62.2	59.1	68.5	62.6	79.9	71.2	65.7
% below cut-off	4.5	7.2	7.2	4.5	6.3	7.2	5.4	6.3
Cut-off % Peretz	73.3	73.3	70	76.7	66.7	73.3	76.6	72.2
% below cut-off Peretz	14.4	22.5	15.3	20.7	7.2	0	9.01	13.5

Table 2.7: Sum of correct responses for the group tested in the laboratory. Cut-off scores and resulting percentage of amusics based on our mean compared to Peretz et al.'s means.

Based on the average of all subtests, 5.4% of the laboratory-tested participants would be diagnosed as amusics because their scores fall below a cut-off score of 71.2% (our mean - 2 SD). The same calculation for the web-based group yields a cut-off score of 55.6% and a categorization as amusic of 6.7%. If the cut-off score by Peretz et al. (2003) were applied, 9% our laboratory-tested participants and 34.6% of the web-tested participants would be categorized as amusic. However, the prevalence is different when considering the subtests individually, now only based on the laboratory group: If only individuals who fell below the cut-off score on every subtest are considered amusic, then the prevalence with the cut-off scores based on our data sinks to 0, and with Peretz et al.'s (2003) cut-off scores it sinks to 4.5%.

Sum of correct responses	**Scale**	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**	**Average**
Mean	24.97	23.86	23.21	24.87	24.09	26.34	24.56
SD	3.03	3.38	3.89	3.80	5.29	3.09	3.94
Cut-off (2 SD)	18.91	17.1	15.43	17.27	13.51	20.16	16.68
Cut-off %	63.03	57.00	51.43	57.57	45.03	67.20	55.60
% below cut-off	5.1	5.1	4.3	8.5	6	6	6.7
Cut-off % Peretz	73.3	73.3	70	76.7	66.7	73.3	76.6
% below cut-off Peretz	20.5	34.2	29.9	26.5	37.6	14.5	34.6

Table 2.8: Sum of correct responses for the web-based group. Cut-off scores and resulting percentage of amusics based on our mean compared to Peretz et al.'s means.

3.3. Subtests

We were further interested in an average score for all three pitch-based subtests, as this is often used in the literature. The following analysis is solely based on the data from the group tested under laboratory conditions due to the very high and improbable percentage yielded by the web-based test. In our sample, the average cut-off score of the pitch-based subtests is 65.7%, yielding 6.3%, (in absolute numbers 7) amusics, while Peretz et al.'s (2003) cut-off scores give a prevalence of 13.5%. We also investigated how many subtests contributed to the pitch average score per subject: Of the 7 amusics below our cut-off score, one fell below the individual cut-off scores on all three subtests, 4 fell below on two subtests and 2 fell below the cut-off score on only one subtest. We then considered again all participants that failed at least one of the three pitch-based subtests, i.e. not only the pitch average, which yielded a total of 13.5% or 15 individuals. It is to be noted that these are not the same 15 individuals that are categorized as amusic when using Peretz et al.'s pitch average cut-off score. The same analysis based on Peretz et al.'s (2003) cut-off scores yielded 26.7% who fell below the individual cut-off scores on all three subtests; 60% fell below on two subtests and 13.3% fell below the cut-off score on only one subtest. Again, when considering all individuals who fell below the cut-off score on at least one of the three pitch-based subtests, 35% (39 individuals) appear to be affected. It is to be noted, however, that a correlation analysis of the scores of the different subtests yielded no statistical reason to use an average score of the three pitch-based subtests. Their scores correlated just as highly with the temporal subtests and the memory subtest as with each other. The scores on the contour subtest, for example, are highly significantly positively related to the scores on all other subtests (contour and scale. $\tau = 0.233$, $p < 0.001$; and interval $\tau = 0.443$, $p < 0.001$; and rhythm $\tau = 0.273$, $p < 0.001$; and meter $\tau = 0.249$, $p < 0.001$; and memory $\tau = 0.199$, $p < 0.001$). A pitch average score can therefore only be motivated by the same component that is supposed to be tested by all three pitch-based subtests but not by a correlation between the scores on these subtests.

3.4. Scoring with Signal Detection Theory

A further analysis was carried out using signal detection theory, in order to inspect whether the obtained scores are tainted by response bias. Therefore, the means and standard deviations of *d’* and *c* were calculated for every subtest, cf. Table 2.9 and Figure 2.2 for the laboratory group. An analysis of skew and kurtosis of *d’* showed that the scores on all subtests are normally distributed.

The previous categorization based on our cut-off scores was kept for this analysis: The group which scored below our cut-off score was labeled “amusic”, while the group that scored below Peretz et al.’s cut-off scores (cf. Table 2.7) was labeled “amusic Peretz et al.”. Our amusic group was a subgroup of the amusic group with Peretz et al. cut-off score for all of the subtests except for the memory subtest, where our cut-off score was higher, cf. Table 2.7. The rest of the participants were labeled “controls”.

The lower part of Figure 2.2 shows the sensitivity measure *d’*, the groups’ ability to discriminate, for each subtest. As can be seen, there is no overlap between our amusic group and the control group for the first five subtests, showing a clear distinction in discriminatory ability between the groups. The *d’* values for the amusic group(s) are much lower than that for the controls for these five subtests, indicating that amusics have difficulties discriminating between the stimuli.

New cut-off scores based on the discriminatory ability of the groups were calculated. The cut-off score was set to be mean - 1 SD (chosen a priori). It is to be noted that it is an arbitrary statistical criterion. Even though the categorization varies based on the statistical criterion that is applied, this might offer a more reliable measure than averaging the sum of correct responses as the bias is factored out and participants can be categorized solely based on their ability to discriminate. The new cut-off scores and prevalences can be found in Table 2.9.

		Scale	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**	**Average**	**Pitch Average**
c	Mean	0.01	0.15	0.29	-0.05	-0.22	-0.21	-0.00	0.15
	SD	0.55	0.54	0.48	0.55	0.34	0.43	0.29	0.46
d'	Mean	2.33	2.25	2.16	2.65	2.81	3.23	2.25	2.25
	SD	0.84	0.90	0.95	0.90	1.28	0.81	0.66	0.73
	z skew	-0.51	0.69	0.28	-0.19	-0.99	0.32	1.56	-0.01
	z kurtosis	1.28	-1.30	0.10	-1.64	-1.70	-1.15	0.61	0.15
% below cut-off score: Mean - 1 SD		15.32	18.02	13.51	25.23	18.92	17.12	14.41	14.41
% overdignosis: Mean - 1 SD		1.80	4.50	1.80	2.70	0.00	1.80	2.70	1.80
% underdiagnosis: Mean - 1 SD		2.70	0.00	0.00	7.21	11.71	11.71	1.80	2.70

Table 2.9: Means and standard deviations of *d'* and *c* for the group tested in the laboratory, including cut-off scores and percentage of amusics and controls categorization based on PC and z-scores used for normality analysis.

This new categorization based on discriminatory ability shows cases of over- and underdiagnosis in comparison to the previous scores. An underdiagnosis (previously categorized as control, but low discriminatory ability) does not happen for two of the three pitch-based subtests. For the scale subtest, it happens for 2.7% of all diagnosis. For the temporal subtests and the memory subtests however, 7.2–11.2% of all participants with a low discriminatory ability are not diagnosed as amusic. Depending on the subtest, an overdiagnosis (diagnosed as amusic, but normal discrimination ability with a high bias) seems to happen in 1.8–4.5% of all diagnosis, based on the entire group. But when only considering the amusia-diagnoses based on the previous scores, then the percentage of overdiagnosis rises to 12–20% depending on the subtest, or even to 30%, when the diagnosis was based on the average score.

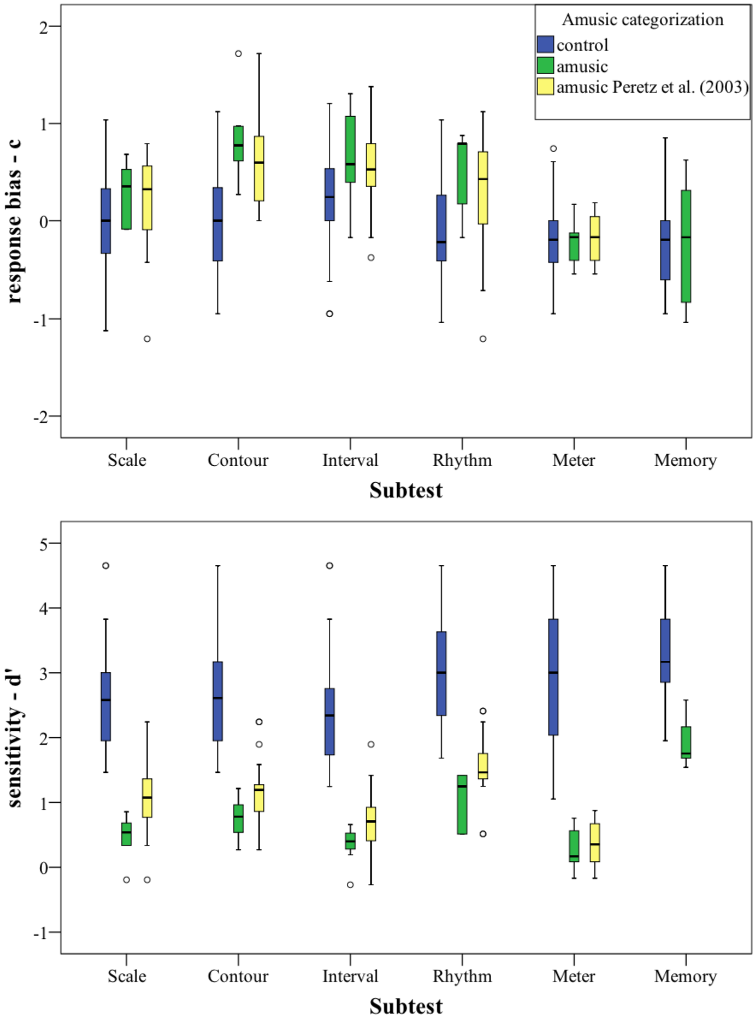


Figure 2.2: Signal Detection Theory scores (*d’* and *c*) plotted per subtest. Categorization based on PC-scores. For statistics, cf. Table 2.9.

In addition, we also calculated the signal detection measures for our web-based data. Table 2.10 gives the means and standard deviations of *d*' and *c* for the web-based group. The analysis of skew and kurtosis of *d'* showed that the mean scores on the scale, contour, interval and rhythm test are normally distributed, while meter and average mean scores exhibit a significant ($p < 0.05$) negative kurtosis value and the memory subtest exhibits a highly significant skew ($p < 0.001$). The cut-off score was once again set to be mean - 1 SD, see last row in Table 2.10.

		Scale	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**	**Average**
c	Mean corr. responses	-0.07	0.17	0.39	0.09	-0.11	-0.27	0.03
	SD	0.52	0.50	0.58	0.55	0.33	0.46	0.54
d'	Mean corr. responses	2.34	2.00	1.95	2.40	2.27	2.88	2.31
	SD	0.93	0.93	1.04	1.12	1.59	1.03	1.17
	z skew	-0.68	1.58	0.66	-0.56	-0.07	***-2.89***	-0.51
	z kurtosis	-1.13	-0.55	-0.65	-0.81	*-2.28*	-0.16	*-2.54*
% below cut-off (Mean – 1 SD)		17.90	12.80	17.10	15.40	15.40	17.10	16.50
% below cut-off (Mean – 2 SD)		1.70	2.60	1.70	1.70	0.90	2.60	1.40

Table 2.10: Means and SD of *c* and *d'* for the web-based group, including cut-off scores and z-scores used for normality analysis. Bold indicates $p < .001$, italics $p < .05$.

3.5. Subtypes

We were also interested in different subgroups, i.e. subtypes of congenital amusia, therefore we considered the different patterns that participants exhibit on the different subtests. This analysis is again solely based on the data from the group tested under laboratory conditions. 53.2% scored below a cut-off score on at least one subtest, 28.8% on at least two and 13.5% on three or more subtests. For the latter two groups, we analyzed the different subtypes. As Table 2.11 shows, there are three distinct subgroups: One bigger group with below cut-off-scores on pitch and rhythm (and partly also memory) subtests and two smaller groups with low scores on only pitch and memory or only rhythm and memory subtests. As many studies also consider the average of all subtests and the pitch-average, we also calculated these. Of the 28.8%, 34.4% also had a below cut-off score on the average of all subtests and on 43.8% also scored below cut-off on the pitch average. For the population with below cut-off scores on at least three subtests, 66.6% had a below cut-off score on the average score and 80% on the pitch average score.

	Two or more subtests		**Three or more subtests**	
	% total	% of below cut-off score	% total	% of below cut-off score
Total	28.8		13.5	
Only pitch	5.4 (3.6)	18.6 (15.6)	2.7 (2.7)	20.0 (20.0)
Only rhythm	9.0 (4.5)	31.3 (12.5)	0.9 (0.9)	6.6 (6.6)
Pitch and rhythm	14.4 (5.4)	50.0 (18.6)	9.9 (5.4)	73.3 (40)

Table 2.11: Percentage of participants scoring below a cut-off score of (mean - 1 SD) on at least two subtests. The value in brackets indicates how many percent also included a below cut-off score on the memory subtest.

3.6. Questionnaire

Our questionnaire contained 27 items: Six demographic items (age, gender, education, handedness, occupation, native language(s)), 20 self-rate items about music education, attitude towards music, music habits and dancing, and one free text question (why people considered themselves unmusical, if they indicated so in the previous question). The questionnaires of the web-based and the laboratory-based group were analyzed together as they were collected under the same circumstances in the laboratory.

Questionnaire item	Component					
	1	2	3	4	5	6
Perception 1 – Melodies without lyrics	**0.85**	-0.16	-0.19	-0.04	-0.07	0.10
Perception 3 – Piano tones	**0.84**	-0.00	-0.08	-0.01	0.02	-0.02
Perception 2 – Off/Wrong singing	**0.72**	0.01	-0.02	0.15	0.07	-0.06
Clapping	**0.69**	0.13	0.20	-0.22	-0.09	0.17
Singing 3 – Notice wrong singing and correct it	**0.52**	-0.07	0.10	0.37	0.06	-0.08
Perception 4 – Surrounded by music as child	**0.44**	0.19	0.12	0.19	-0.03	-0.12
Music Education 2 – Age of onset	0.01	**0.90**	-0.08	-0.05	-0.09	-0.02
Music Education 4 – Frustration	0.00	**0.86**	0.00	-0.04	0.00	0.03
Music Education 3 – Years of lessons	-0.14	**0.82**	0.09	-0.03	-0.09	0.06
Music Education 1 – Type of education	0.15	**0.60**	0.00	0.06	0.02	-0.12
Music Education 5 – Still playing/singing	-0.08	**0.52**	-0.16	0.19	0.32	0.05
Dancing 2 – Quality – Own assessment	0.03	-0.01	**0.95**	-0.06	0.00	-0.05
Dancing 1 – Quantity/Frequency	-0.14	-0.03	**0.91**	0.13	0.04	0.04
Singing 1 – When alone	-0.08	-0.01	0.09	**0.85**	-0.04	0.13
Singing 2 – In public	0.13	0.01	-0.05	**0.80**	-0.02	-0.01
Unmusicality 1 – Family members	-0.15	-0.11	-0.01	0.04	**0.93**	-0.01
Unmusicality 2 – Own assessment	0.17	0.04	0.08	-0.13	**0.83**	0.05
Listening habits 1 – Quantity listening to music	-0.02	-0.02	-0.03	0.27	-0.13	**0.80**
Listening habits 2 – Attitude towards music	0.09	0.02	0.02	-0.12	0.17	**0.77**

Table 2.12: Summary of principal component analysis: Rotated pattern matrix with factor loadings, ordered according to the factor loadings per component.

A principal component analysis (PCA) with an oblique rotation (promax) allowed for the collapsing of the items into 6 factors. The PCA included 19 out of the 20 self-rate items, as the remaining item (Perception 5 – Evaluation of own perception) failed to reach the acceptable limit of 0.5 on the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy. This item was included in a first analysis but excluded from the final analysis. It is to be noted that not every participant answered every question and cases were therefore excluded pairwise. The KMO measure verified the sampling adequacy for the analysis, KMO = 0.8, and all KMO values for individual items were > 0.5. Bartlett's test of sphericity χ^2 (171) = 1530.778, $p < 0.001$, indicated that correlations between items were sufficiently large for PCA. Six components had eigenvalues over Kaiser's criterion of 1 and in combination explained 67.65% of the variance. Table 2.12 shows the factor loadings after rotation. The items that cluster on the same components suggest that component 1 represents *perception* but also contains clapping, component 2 *music education*, component 3 *dancing*, component 4 *singing/production,* component 5 *self-assessment of musicality* and component 6 *music listening habits*. These components were then entered into a multiple regression analysis in order to calculate their influence on the MBEA-scores.

The multiple regression was performed separately for every subtest. The regression analysis again included only the MBEA-scores obtained in the laboratory-based group and only those participants who answered all questionnaire items, so as not to include participants with missing values. 76 participants remained in this analysis step. The six components were entered as predictors and *d'* was used as outcome variable. To use *d'*-scores in this context has an advantage over using PC scores, as these were shown not be normally distributed, which is one of the assumptions that has to be met for a regression analysis. The assumptions of multicollinearity and independent errors were true (cf. Table 2.13 collinearity statistics and Durbin-Watson test respectively). The assumptions of linearity and homoscedasticity were visually inspected and also true.

Different models were fit to the data, excluding non-significant predictors, until the best fitting regression model was found for every subtest. Table 2.13 summarizes these models.

		Coefficients		Collinearity Statistics		Model Fit		
Subtest	Components included	B	β	VIF	Tolerance	R^2	Durbin-Watson	ANOVA F-ratio
Scale						***0.22***	1.87	***10.23***
	Constant	***2.33***						
	1 perception	*0.20*	0.23	0.83	1.20			
	2 music edu.	*0.27*	0.33	0.83	1.20			
Contour						***0.22***	2.41	***10.12***
	Constant	***2.20***						
	1 perception	***0.29***	0.31	0.83	1.20			
	2 music edu.	*0.21*	0.24	0.83	1.20			
Interval						***0.25***	1.93	***24.70***
	Constant	***2.01***						
	2 music edu.	***0.46***	0.50	1.00	1.00			
Rhythm						*0.09*	2.07	*6.90*
	Constant	***2.75***						
	2 music edu.	*0.24*	0.29	1.00	1.00			
Meter						***0.38***	2.13	***22.30***
	Constant	***2.76***						
	2 music edu.	***0.73***	0.56	0.99	1.01			
	6 attitude	*0.25*	0.21	0.99	1.01			
Memory						*0.11*	2.10	*8.97*
	Constant	***2.93***						
	5 own musicality assessment	*0.31*	0.33	1.00	1.00			
Average						***0.39***	2.36	***23.30***
	Constant	***2.22***						
	1 perception	*0.21*	0.30	0.83	1.20			
	2 music edu.	***0.30***	0.44	0.83	1.20			

Table 2.13: Summary of multiple regression analysis predicting MBEA scores per subtest from components. Bold and italics indicates $p < 0.001$, italics $p < 0.05$.

R^2 can be used as a measure of how much variability of the outcome variable is accounted for by the predictors. 22% to 25% of the variation in MBEA scores of the three pitch-based subtest can be predicted by the first and second component. Only 9% of the variation of the rhythm scores can be explained by questionnaire items, while 38% can be explained for the meter subtest, based on component 2 music education and component 6 listening habits/attitude. 11% of the variability in memory scores can also be explained by questionnaire items, more specifically by one's own musicality assessment. Lastly, the 39% of the variation in the average score can be predicted by component 1 and 2, perception and music education respectively. The standardized beta value (in standard deviation units) is used as a measure of how much the outcome variable is changed by a change of the predictor. The standardized beta value on the interval subtest (standardized $\beta = .50$), for example, indicates that if the score of component 2 increases by 1 standard deviation (1.02), the *d'* score increases by 0.50 standard deviations (0.47).

4. Discussion

In the current paper, we analyzed a sample of 228 German undergraduate students who were assessed with the MBEA and a comprehensive questionnaire. We compared the differences between scores that were obtained in a web-based study (at participants' homes, N = 117)) and those obtained under laboratory conditions with a computerized version of the MBEA (N = 111). In addition to traditional statistical procedures, the data were evaluated using Signal Detection Theory (Green and Swets, 1966). We investigated the prevalence of congenital amusia, subtypes and the usage of different subtests.

Before looking in detail at our results, we have to note that, like in the studies by Wise (2009), the MBEA cut-off scores proposed by Peretz et al. (2003) yielded a very high and improbable percentage of amusics (34.6%) for our web-based group. We therefore used the cut-off scores calculated on the basis of our own data.

4.1. Scoring, Prevalence and Misdiagnosis

In the present study, a comparison of the MBEA scores for our laboratory-tested participants calculated both on the basis of the sum of correct answers and the Signal Detection measure *d'* (with mean - 1SD as cut-off) yielded different diagnoses. With the PC-based measure, 12–20% of amusia diagnoses are misdiagnoses of people who could be shown to have a normal discriminatory ability but simply a larger response bias. If we consider the average score across all subtests, then this misdiagnosis rises to 30%. This number confirms Henry and McAuley's (2013) finding of a PC-based misdiagnosis of 33%. At the same time, the PC-based scores in our study fail to diagnose 7.2–11.2% of the participants with a low discriminatory ability as amusic.

Furthermore, we found that 28.8% of the laboratory-tested participants scored below cut-off score on two or more subtests and of those only 34.4% also scored below cut-off on the average of all subtests. This shows that a substantial number of participants with an impaired discriminatory ability is missed by using average scores for the diagnosis of congenital amusia. In addition, we could show that for the average pitch score, which is often employed in MBEA studies (see the overview in section 1.2.1), the scores on the pitch subtests correlated as highly with each other as with the other subtests, giving no statistical reason for using an average pitch score.

The misdiagnosis of congenital amusia has implications for the inclusion of participants in scientific studies and therefore the expansion of knowledge about congenital amusia. At the same time, the diagnosis has personal consequences for the individual in question, just as in the case of acquired amusia. Many possible amusics who come to our lab actively seek answers as to why their perception seems to be different from that of other people. These participants deserve an accurate assessment of their abilities. Using *d'* to assess amusics' discriminatory ability reflects their abilities more accurately than using the sum of correct answers.

4.2. Subtypes

We were furthermore also interested whether our data provided evidence for different subtypes of amusia, as have been proposed in previous studies (see the discussion in section 1.2.5). For our group of participants that performed below the cut-off score of mean minus 1 SD and failed at least two subtests, we found three subgroups: A group that only exhibits pitch deficits (18.6% of amusics), one with only rhythmic deficits (31.4% amusics) and another with pitch and rhythm deficits (50% of amusics). All of these groups contained participants with and without low performance on the memory subtest. When considering only participants who failed at least three subtests, then the same three groups remain. However, only 6.6% of amusics exhibit a rhythm deficit, 20% exhibit a pitch perception deficit and 73.3% exhibit both. The probability of these three types is the same as if the three failed tests were randomly distributed across all six subtests. The high co-occurrence of pitch and rhythm deficits could be due to the very high correlations between the various subtests, which were found above. The percentage of the rhythm subtype sinks so drastically due to the imbalance of pitch and rhythm tests on the MBEA. In order to score below cut-off on three subtests with only one rhythm perception deficit is impossible, therefore also a memory deficit has to be present. Reports of subtypes with pitch deficits or pitch and rhythm deficits are more frequent than cases with rhythmic problems only (Phillips-Silver et al., 2011: 1 case; Launay et al., 2014: 3 cases), possibly due to the low proportion of rhythm-related subtests in the MBEA. We therefore propose that additional tests assessing rhythmic abilities, e.g. a part of the Beat Alignment Test by Iversen and Patel (2008), should be considered as a supplement to the MBEA. This might make a further differentiation of subtypes of congenital amusia and a clearer definition of them more feasible in the future. This finding again also supports our view that an average score should not be used for the diagnosis of amusia (see also Wise (2009) and Henry and McAuley (2010)), as it does not reflect the heterogeneous behavior of participants across the six subtests. The evaluation of scores on individual subtests, on the other hand, can lead to misdiagnosis of people as amusic who simply did not pay enough attention to the experiment. Though we tried to filter out such participants by the so-called catch trials, these catch trials can be detected without focused attention and therefore might not be an adequate way of controlling for such possible false positive diagnoses.

4.3. Questionnaires

In addition to the MBEA scores, we also analyzed the information from our questionnaire. The questionnaire contained 27 items, which were reduced by principal component analysis to 6 components. These encompassed perception, music education, dancing, singing/ production, listening habits and self-assessment of musicality. Three of these components, singing/production, music education and listening habits, overlap with the ones found by Cuddy et al. (2005). She identified a total of four components, the fourth being childhood memories. While Cuddy et al. (2005) and Wise (2009) were interested in the self-labeling as tone-deaf and consequently used it as outcome variable, we incorporated it as one of our components, 5 - self-assessment of musicality, into a multiple regression analysis with *d'* as the outcome variable, because both studies found self-reports of tone-deafness not to be reliable and overlapping with the presence of congenital amusia. Our analysis showed that part of the variation in *d'* scores can be accounted for by questionnaire information, more specifically music education and perception. The outcome of the meter subtest was also influenced by listening habits, a finding that is in agreement with Cuddy *el al.*'s (2005) findings. Contrary to Cuddy et al., we found no influence of the component music production on our outcome variable. We also found that only (and also only a small but significant amount of) the variation of scores on the memory subtest can be accounted for by one's own assessment of musicality. It did not account for any other variability in *d'* scores. Considering these results, it seems adequate to use at least a short questionnaire containing items about music education and music perception.

4.4. Web-based versus laboratory testing with the MBEA

For the scoring based on the sum of correct responses, a slightly higher proportion of web-tested participants fell below the cut-off score and thus was diagnosed as amusic (6.7%) than for the group tested in the lab (5.4%). This contrasts with the findings by Peretz et al. (2008) who report that 19% of their participants were diagnosed as amusic in the laboratory would have been missed as such by their online test. As explained by Peretz et al., their findings are due to a difference in task: Whereas for the on-line test participants had to spot possible incongruities in melodies, for the lab-used MBEA they had to compare two melodies at a time, which is more demanding as it requires the storage of pitch information in the working memory. For the present comparison between lab and web-based testing we used the same test (the full MBEA), hence the differences we found have to be attributed to the testing method.

When looking at the differences in performance in the individual subtests, we found that the scoring based on the sum of correct responses yielded non-normally distributed results that are highly negatively skewed for all subtests. This is in accordance with the findings by Wise (2009) and Henry & McAuley (2010), and led us to use non-parametric statistics for the comparison of the two groups. Mann-Whitney-U tests revealed significant differences between the groups on the contour, interval, meter and memory subtests as well as on the average score, with the performance on the web-based version being worse.

For the further analysis of the scores we then employed the signal detection theory (SDT) measures *d'* and *c* (as suggested by Henry & McAuley, 2013). For the lab-based sample, the *d'* scores for all subtests and the average score were all distributed normally, indicating that the discriminatory ability of this group was fairly consistent. For the web-based group, only four of the six subtests were distributed normally. The meter subtest exhibits a significant kurtosis value, while the distribution of scores on the memory subtest is highly significantly negatively skewed and platykurtic, i.e. it contains many high scores but also exhibits a long-drawn tail to the left with low scores and an overall flat distribution. For these two subtests, the web-based group thus shows less discriminatory abilities. Possible explanations for this difference in discriminatory power and also for the statistical difference in correct scores between the web-tested and the lab-tested groups for four of the subtests are discussed below in sections 4.4.1. on MBEA-related issues and section 4.4.2. on web-based testing in general.

4.4.1. The MBEA as a web-based test and its limitations

The MBEA was not designed for web-testing, and some of its properties do not seem to make it ideally suited as a web-based test for amusia. In this subsection, we discuss two of these properties, namely length and lack of measures to ensure participants' attention as possible reasons for the low performance of our participants on the last two subtests of the MBEA.

With respect to length, the whole MBEA takes 50 minutes on average to complete under laboratory conditions. While the majority of our participants also completed the web-based version within 50-60 minutes, some took over 90 minutes. This time seems to be too long for a web-based study, as the literature shows. Reips (2002) suggests "a few minutes", Honing & Ladinig (2008) 15 minutes, and Gingras et al. (2015) 30 minutes as preferred length for web tests. While Peretz et al. (2008) used a shorter web test loosely based on the MBEA, this yielded misdiagnoses in both directions, as discussed above. Some studies (e.g. Liu et al., 2010, 2013; Williamson and Stewart, 2010) only use two subtests of the MBEA for pretesting participants via the web, as mentioned in the introduction.

In order to ensure the participants' attention, the MBEA contains four catch trails (as described in the methods section). All of our participants that finished the web-based test had scored correctly on all four catch trials. However, it is a relatively low number of catch trials and the manipulation in the catch trails stands out so much from the experimental manipulation that anyone paying only the slightest bit of attention should be able to identify them. The catch trials are therefore not enough to ensure a participants' lasting attention, especially in web-testing, where a very quiet and non- distracting environment cannot be assured.

Both factors, length of test and inability to ensure participants' attention, could thus have led to a worse performance of the web-tested group. Especially lack of attention during web testing could contribute to the lower and non-normally distributed discriminatory ability on the last two subtests but especially on the memory subtest as this test relies on how much participants paid attention to and remember from the first four subtests.

4.4.2. General limitations of (auditory) web-based testing

There are a number of advantages as well as limitations of web-based testing that are not specific to the MBEA but apply to all, or at least all auditory, web-based studies. Especially the limitations will be discussed here (as possible explanation of our findings). These issues are not new and have been raised before (see for example Mehler, 1999; Krantz & Dalal, 2000; Reips, 2002; Birnbaum, 2004; and a discussion on the auditory mailing list: Auditory 2007) but we will discuss them in light of our experiences with the MBEA. Where solutions have been proposed in the literature (e.g. Reips, 2002; Birnbaum, 2004), they are also outlined.

The obvious **advantages** of web-based testing are that it is relatively easy to gather large heterogeneous data samples as well as to reach specialized populations easily. This can be achieved at much lower costs and at a higher speed than in traditional laboratory settings, while offering a greater external validity and using more automated processes, which makes data analysis faster as well (Reips, 2002). Furthermore, web-based studies using highly standardized procedures are easily replicated (Birnbaum, 2004) and provide a much more natural listening setting (Honing & Ladinig, 2008) that avoids an experimenter bias, i.e. participants do not feel pressured to respond in one way or the other due to the presence of the experimenter.

Though motivation to take part in a study is usually an advantage of web-based testing, our participants had to take part in the study for course credit and therefore were not as intrinsically motivated to participate as other subjects in music-related studies (see e.g. Honing and Ladinig, 2008 on their very positive experience with musically interested participants in music-related web testing). However, our laboratory participants also had to take part for course credit, therefore a low motivation cannot explain the difference in performance between the two groups.

The most prominent **disadvantages** of web-based testing in general are the high drop-out rates and multiple submission, both of which are threats to the internal validity of studies (c.f. Reips, 2002; Birnbaum, 2004).

Generally, a **drop-out rate** of 30–40% has been reported for web-based studies (Reips, 2002). In our study, we observed a dropout rate of only 10.7%, which can be attributed to the fact that students had to participate in the study for course credit. In addition, we assured our participants of confidential handling of the data before the study. We also made them aware of the possibility of back-tracing data to participants, and the availability of an explanation of the aim of the experiment after participation, thereby showing that the data were actually analyzed and used in a scientific context. This latter fact greatly interested at least part of the students and many of them not only wanted to be informed about the general outcome of the study but also about their personal results.

Reips (2002) proposes the use of the so-called high hurdle technique against high drop-out rates. With this technique a web-study is designed in such a way that ‘obstacles’ that test participants’ patience are put at the beginning, e.g. the collection of personal data or a screen that takes long to load. With this, it is hoped that all impatient participants or participants that are unwilling to provide personal information are filtered out before the beginning of the actual study. This method not necessarily reduces the drop-out rate but ensures that uninterested participants drop out as early as possible thereby avoiding incomplete datasets. Other measures against a high drop-out rate can be the promise of rewards or a design of the test that is visually appealing or intellectually challenging, as is recommended by Honing & Ladinig (2008) However, as Reips (2002) points out, experiments that are too interesting or engaging might provoke **multiple submissions**. Our MBEA-online version was designed in such a way that it exactly mirrored the visually rather plain instruction screens that were used for the computerized laboratory version of the MBEA implemented with Praat (Boersma & Weenink, 2011), again not tempting our participants to perform the test several times. Reips (2002) makes several suggestions for the avoidance or the control of such multiple submissions, such as the collection of personal data for identification, the tracking of IP-addresses, the implementation of a username and password-dependent access. All of these were implemented in our study and we did not have a single multiple submission. The same is observed by Birnbaum (2004) who found only one instance of multiple submission in a dataset of 1000 submissions.

A further disadvantage of web-testing is the **lack of control** pertaining to technical factors (e.g. internet speed or usage of headphones) and environmental factors (like noise or distractions). Both can influence the data considerably (Mehler, 1999; Auditory 2007). It has been argued that lack of control is not actually an issue and that web-based studies have a much greater external validity (Reips, 2002; Gingras et al., 2015) through their large participant numbers that cancel out the possible noise in the data (McGraw et al., 2000). However, Krantz and Dalal (2000) argue that for auditory research, a stable and quiet environment is crucial for the success of the experiment. Auditory (2007) shows that many researchers are in doubt about the use of web-based experiments in auditory research, since it cannot even be controlled whether subjects wear headphones or what the level of background noise is during the experiment.

Concerning internet connection and speed, Reips (2002) advises to pre-load all soundfiles. This takes longer at the beginning of the experiment but ensures then that the experiment can start and run smoothly. To ensure smooth running, experiments should be checked beforehand on different operating systems and with different browsers (Reips, 2002). In our web-testing, we followed these recommendations, but nevertheless 23 of our participants encountered technical difficulties. These could mostly be resolved by updating browser versions or installing or updating plugins. However, this factor might have influenced the difference in performance between the two groups in our experiment.

Further issues on lack of control that clearly influenced our results are the environmental factors. We instructed all participants to wear headphones (not to use their computer speakers) and to take the experiment in a quiet room without distractions or interruptions. Whether they followed these instructions could not be checked. Furthermore, they were asked to finish the experiment in one instance and to only take a break when they needed one. As we logged the time of day during which participants took the test and how long they took for every subtest, we could see that data was submitted round-the-clock and that some participants took very long to finish certain blocks, not all did thus follow our instructions. A possible way around this last problem could be the exclusion of participants on the basis of such long times. The long breaks taken by some participants could contribute to the lower and non-normally distributed discriminatory ability especially on the memory subtest.

Lack of control of the environmental factors on the site of the experimenter is thus a crucial point that makes web-based testing unsuitable for the MBEA or for more than just a pre-screening.

A general point of concern with web-based testing not connected to performance is that **of security/privacy concerns**. Via http protocols or Javascript it is possible to track sensitive information about the participant: Which operating system and browser are used, screen resolution, loading times, which link referred them, and even the location can be tracked and logged. Participants need to be informed what data is or even can be collected about them and how it is being stored. However, this information is often not provided, which raises ethical concerns. Indeed, many ethics committees do not approve web-based studies and some journals will not accept web-based studies for publication (Auditory 2007; Honing & Ladinig, 2008).

One last concern that is also related to ethics is important to consider when screening for amusia online, be it with the MBEA or any other kind of test, namely that of **diagnosis**. Most participants that will voluntarily seek out a web-based amusia test do this because they suspect that “there is something wrong” with them, i.e. that they have a perceptual deficit and are amusic. These participants naturally take a test like that to get to know their results on that test. However, it is questionable whether and how these participants should be informed of their results. In the case of the present study this was comparatively simple. Participants did not automatically receive their results. Pooled results were presented to all participants. More interested participants could request their personal results, which they were then given including a detailed explanation of what these results meant or did not mean. However, this is not possible with large online samples stemming from the general public. It is questionable whether a relatively simple web-based test should "diagnose" people with a life-long affliction. In a lot of cases when amusics were diagnosed in our laboratory, they were glad when they learned of their amusia because they finally knew "what was going on" with them. However, in a few cases it was almost traumatic for people and these people should not be confronted with a diagnosis like that while sitting alone in front of a computer screen without further explanation.

Finally, this last example also nicely exemplifies another issue of web-based testing: **Self-selection**. While it is argued that a more heterogeneous pool of participants can be reached via internet – and this is certainly true if compared to the normal psychology undergraduate participant pool – the sample one obtains might still be biased. Only people very interested in their own musicality or people doubting their musicality will actively seek out web-based musicality or amusia tests. This also yields a participant pool that does not reflect the normal population but rather two extremes.

While web-based testing thus offers many advantages and is suitable for many kinds of research, we would like to caution that it might not be suitable for the diagnosis of amusia (with the MBEA or another test). Web-based tests can certainly be used as pre-screening tools (as parts of the MBEA are used at the moment) and can be very useful as such. But for the various reasons outlined above, an amusia diagnosis, even if it is no medically recognized diagnosis, should not take place via a web-based test.

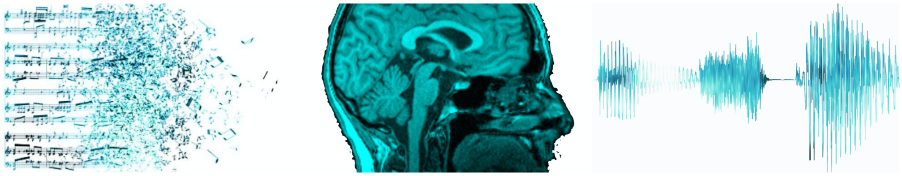
Concerning the MBEA, we showed that even though the sum of correct responses differed significantly between our web-tested and laboratory-tested groups, their discriminatory ability was relatively similar. Only the last two subtests showed differences between the two groups, but these can probably be attributed to some properties of the MBEA that make it in its entirety unsuitable for online testing.

5. Conclusion

In sum, we thus recommend the calculation of cut-off scores based on the SDT measures *d'* and *c* instead of percentage correct for all MBEA subtests separately (rather than averaging over subtests) and the additional use of a questionnaire and a further rhythmic subtest. We furthermore advise testing in the laboratory only. This way, a more reliable diagnosis of congenital amusia and a differentiation of amusic subtypes seem possible in the future.

6. Acknowledgements

We would like to thank Henk Schotel for his help with the web-based implementation of the MBEA.



3. Vowel Perception in Congenital Amusia

<u>This chapter is yet to be submitted as:</u>

Pfeifer, Jasmin & Silke Hamann. Vowel Perception in Congenital Amusia

Abstract

Congenital amusia is a disorder that negatively influences pitch and rhythm perception and it is not caused by a hearing deficiency or brain damage. While congenital amusia had long been reported to affect only the musical domain (Peretz et al., 2002; Ayotte et al., 2002), several studies have shown that amusics also have impaired perception of intonation (e.g. Patel et al., 2008) and linguistic tones (e.g. Tillmann et al., 2011).

In the present study we tested whether congenital amusia also has an influence on linguistically relevant cues other than pitch by investigating the discrimination of German front mid vowels /ɛ/, /ɛ:/, /e/ and /e:/ (where /ɛ:/ and /e/ are of low frequency, the latter being a loan phoneme). We assessed amusics' behavioral responses with an AXB task and their electrophysiological responses, more specifically the Mismatch negativity (MMN), which is a component evoked by unconscious change detection in the auditory signal (e.g. Näätänen, 2001), with a multi-deviant oddball paradigm in four blocks.

We tested 11 congenital amusics diagnosed with the MBEA and 11 matched controls. Our stimuli were isolated synthetic vowels created by Klatt synthesis, varying in either durational or spectral properties (F1 and F2 varied together), which resulted in four continua with seven steps.

In the behavioral study, amusics performed worse than controls and a difference in duration was overall for both groups harder to detect than a difference in formant frequency. For the MMN data, we found that amusics exhibit an MMN albeit a significantly reduced one. In addition, we again found that durational differences were harder to detect, especially for amusics.

Our study shows that congenital amusia does not only affect the perception of pitch in music and language but also the perception of vowel contrasts, therefore having more far-reaching consequences for speech perception than previously assumed. Not only was the behavior of amusics shown to be affected, we also found differences in the MMN, reflecting differences in early auditory change detection.

1. Introduction

Congenital amusia (henceforth: Amusia) is an innate disorder that causes lifelong deficits in pitch and partly also rhythm perception. The disorder is not caused by hearing loss, brain damage or insufficient music exposure (Ayotte, Peretz, & Hyde, 2002). The most apparent symptoms to the affected individuals themselves are various inabilities in the musical domain such as: Recognition of familiar melodies, detection of out-of-tune notes or singing, or an inability to clap or sing along. Possibly due to those clear symptoms, early research has mostly been focused on its influence on music. Hence, congenital amusia has long been characterized as a music-specific disorder (Ayotte et al., 2002; Peretz et al., 2002; Peretz, Blood, Penhune, & Zarorre, 2001). Various areas of music perception and musical engagement have been assessed over time and found to be impaired by amusia, such as pitch perception (Peretz et al., 2002), pitch production (Dalla Bella, Berkowska, & Sowiński, 2011), rhythm perception (Foxton, Nandy, & Griffiths, 2006), beat synchronization (Sowiński & Dalla Bella, 2013), timbre perception (Marin, Gingras, & Stewart, 2012), consonance rating (Ayotte et al., 2002), and musical emotion perception (Marin, Thompson, & Stewart, 2012). Several hypotheses have been posited and tested concerning the underlying deficit of amusia. For the longest time, amusia was said to be caused by a fine-grained pitch processing deficit (Ayotte et al., 2002). However, this deficit could not account for various symptoms in amusia. Therefore, further underlying causes were assessed such as a pitch memory deficit (Gosselin, Jolicœur, & Peretz, 2009), a statistical learning deficit (Peretz, Saffran, Schön, & Gosselin, 2012) or a rapid-auditory processing deficit (Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010). There is still no unequivocal consensus concerning the underlying cause and it seems most likely to be a multi-causal deficit that is responsible for the different symptoms exhibited by amusics.

Recently, studies have started to investigate amusics' processing of other parts of the auditory signal than pitch. Cousineau, Oxenham, & Peretz (2015) found that amusics had deficits in the processing of temporal fine structure but not in the processing of the temporal envelope. Interaural time differences, intensity and spectral resolution were also tested and found not to be impaired. Based on this, the authors conclude a pitch-specific deficit in spectro-temporal processing unrelated to temporal or spectral coding in the auditory periphery. Supporting these findings, Whiteford & Oxenham (2017) discovered frequency modulation deficits, possibly causing problems in representation and coding of frequency, as well as amplitude modulation deficits, possibly causing deficits of fine-grained perception other than frequency and pitch. Bones & Wong (2017), on the other hand, found no insensitivity to envelope cues but rather an over-reliance on these cues. However, they also highlight that amusic deficits are not limited to fine-grained pitch processing.

Musical impairments were the focus of initial research on amusia. However in more recent work, the impact of amusia on language perception is being investigated, as pitch is also used as an important cue in it: Pitch is used in intonation to disambiguate questions from statements or to mark focus for example. Intonation perception was also the first area of speech perception to be shown to be affected by amusia (Patel et al., 2008; Liu et al., 2010; Hamann et al., 2012). Pitch is also used in other linguistic areas, which have been shown to be affected as well, such as tone language perception (Liu, Jiang, Wang, Xu, & Patel, 2015; Liu, Maggu, Lau, & Wong, 2015; Liu, Xu, Patel, Francart, & Jiang, 2012; Tillmann et al., 2011a) and emotional prosody in language (Lolli, Lewenstein, Basurto, Winnik, & Loui, 2015; Thompson, Marin, & Stewart, 2012). Due to these findings, congenital amusia is now seen as a domain-general disorder that negatively affects pitch processing in general (Liu et al., 2010; Hamann et al., 2012; Zhang, Shao, & Huang, 2017).

So far, the study of speech perception impairments caused by amusia has almost exclusively focused on areas involving pitch as a perceptual correlate of speech sounds, which is probably due to the fact that most hypotheses on the underlying deficit of amusia are based on some form of pitch perception deficit. Speech, however, also makes extensive use of other information in the speech signal such as spectral frequencies. The latter are especially relevant in the perception of vowels. Vowel spectra have peaks and valleys, which are characterized by increasing or decreasing amplitude of their harmonics (e.g. Hayward, 2000). The frequency bands of these peaks are called formants. The quality of a vowel is usually determined by its lowest two formants, where the first formant (henceforth: F1) correlates with the tongue height in the mouth when producing the vowel (low F1 stands for a high tongue position, and vice versa), while the second formant (henceforth: F2) correlates with the horizontal position of the tongue (low F2 stands for a back position, high F2 for a more frontal position). According to the tongue position, and thus F1 and F2 values, vowels are characterised as high vs. low, and front vs. back. Other aspects, such as nasality, rounding of the lips, and whether a vowel is steady-state or shows inherent movement also influence formant values, but are neglected in the present study as the vowels were chosen accordingly.

First reports on vowel perception by congenital amusics appeared recently, all testing native speakers of tone languages. Huang et al. (2016) assessed the discrimination and identification of two high vowels in Mandarin and found no difference between amusics and controls in the identification task, but the overall discrimination rate of amusics was lower. Zhang et al. (2017) had similar findings for the perception of two back vowels in Cantonese, and concluded that the deficit in amusia is not specific to pitch processing but rather concerns the processing of spectral frequencies in general. Tang et al. (2018) investigated Mandarin amusics' vowel perception (among other things). They divided their amusic group into what they call "pure amusics" and "tone agnosics", the latter displaying speech tone difficulties. Only tone agnosics exhibited impaired vowel perception (and only in one noise condition).

These existing studies on amusic vowel perception only considered the differences in vowel quality and neglected durational differences, though Tang et al.'s (2018) vowel stimuli ranged between 236 ms and 772 ms, which could have served as perceptual cue for amusics, explaining why their performance was comparable to that of the controls. Furthermore, all studies involved tone languages, while the perception of vowels by amusic native speakers of non-tone languages has not been tested yet. And lastly, all three studies employed only behavioral tasks to test vowel perception, while research on speech perception of non-amusics and on pitch perception by amusics showed interesting results on the electrophysiology underlying such behavior.

Different event-related potential (ERP) components, measuring the brain's electrophysiological response to a stimulus, can be used in speech perception research. One such component is the so-called mismatch negativity (MMN), an early component that reflects an automatic, unconscious detection of change in a series of stimuli, which can be recorded without requiring attentive action from the participant (Näätänen, Gaillard, & Mäntysalo, 1978) for reviews, see: Näätänen (2001); Näätänen, Paavilainen, Rinne, & Alho (2007); Picton, Alain, Otten, Ritter, & Achim (2000). The MMN can also be used as an index of behavioral accuracy (e.g. Näätänen et al., 2007): Large MMN amplitudes indicate an accurate stimulus discrimination, while small MMN amplitudes have been shown to be associated with inaccurate discrimination (Kujala & Näätänen, 2001). This holds true for healthy groups as well as various patient populations.

The MMN can be elicited by a change of frequency, duration, intensity, location or pattern in a stream of speech or non-speech sounds (Partanen et al., 2011; Ylinen et al., 2006). Generally, so-called oddball paradigms are used: A repetitive standard stimulus is presented many times in a row and it is occasionally interrupted by a deviant differing in an acoustic feature (Näätänen, 1990; Schröger, 1997). It is also possible to use multiple, different deviants within a string of one standard in so-called multi-deviant oddball paradigms. The MMN typically peaks at around 100 to 250 ms when a deviant is presented, if the participants' auditory system has formed a representation of the repetitive aspect of the standard stimulus (Näätänen, 2001).

The MMN is ideally suited as a starting point in the investigation of electrophysiology of vowel discrimination in amusia, as MMN paradigms have long been used in general auditory but also speech perception research e.g. Chládková, Escudero, & Lipski, 2013; Kirmse et al., 2008; Näätänen, 2001; Partanen, Vainio, Kujala, & Huotilainen, 2011; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006. The linguistic MMN is thought to arise not only from auditory change detection but also from the representation of speech sounds in long term memory that facilitate the discrimination process, e.g. Näätänen, 2001; Partanen et al., 2011. Partanen et al. (2011) have shown that the MMN in linguistic research can be used to establish an auditory discrimination profile, taking into account duration, intensity, pitch and vowel differences. Especially vowel quantity, vowel quality (Chládková et al., 2013; de Jonge & Boersma, 2015) or both taken together (Kirmse et al., 2008; Partanen et al., 2011; Ylinen, Huotilainen, & Näätänen, 2005) have been researched with the MMN. Vowel quantity, contrary to other linguistic features, has been shown to be more right-lateralized (Kirmse et al., 2008; Partanen et al., 2011) and to elicit bigger MMN amplitudes than vowel quality changes (Partanen et al., 2011), whereas simultaneous changes in quantity and quality elicited the biggest amplitude (Ylinen et al., 2005). The MMN has been shown to originate in the auditory cortex and the fronto-central scalp areas (for a review see Näätänen et al. 2007), and in a lesion study specifically the dorsolateral prefrontal cortices (Alain, Woods, & Knight, 1998) were implicated. These are also the regions that are affected by amusia, as shown by Albouy et al., 2013b; Hyde, Zatorre, & Peretz, 2011. The MMN therefore seems well suited to investigate the neurophysiological processes underlying congenital amusia.

In addition, a number of studies investigating pitch perception in amusia have made use of the MMN already. However, there were no uniform findings, they differed rather widely: Braun et al., (2008) found the MMN to be absent in amusics for melodies containing altered notes. On the other hand. Moreau et al., (2009) testing adults with melodies and Mignault et al., (2012), testing children with tonal sequences, found normal MMNs in amusics. The same holds for the study by Moreau et al. (2013), testing amusics with piano tone sequences. Reduced, abnormal MMNs were found by Lebrun et al. (2012) (for one child to small tonal changes of 25 per cent) and by Nan et al. (2016) for the responses of tone-language speaking amusics to lexical tones. The latter found a reduced MMN only for a subgroup of amusics, namely for tone agnosics.

Taken together, the aforementioned findings on amusia show that behaviorally, vowel perception in a sub-group of tone-language amusics seems to be impaired, with no research into the underlying cognitive processes yet. On a broader level, some amusics at least seem to show absent or reduced MMNs to tonal sequences. All studies highlighted that it is important to carefully screen the amusics and to test a group that is as homogeneous as possible in their deficits in order for a clear picture to emerge.

Based on this, the present study looks at a fairly homogeneous group of amusics (showing both a pitch and a rhythm deficit) of a non-tonal language, German, with a contrast in vowel quality and quantity. We hypothesize that these amusics will perform behaviorally worse in discriminating carefully controlled-vowel stimuli and will show reduced MMNs in comparison to controls. We also assess whether the expected deficit is present both in vowel quality and vowel quantity, expecting both will be impaired due to the participants' deficit in pitch and rhythm perception. To test these hypotheses, we designed a behavioral study (with an AXB task) and an electrophysiological study.

2. Materials and Methods

2.1. Participants

11 amusics and 11 controls matched for age, gender, handedness, education and musical training participated in both studies. The participant details are listed in Table 3.1. All participants were native speakers of German, right-handed and had no self-reported psychological or neurological disorders. They had normal hearing defined as a mean hearing level of 20 dB or less in both ears, assessed by a pure tone audiometry at 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz.

Group		**Age**	**Years of education**	**Years of music education**	**Gender**
Amusic	Mean	34.09	19.64	1.64	3 male
Control	Mean	32.45	18.64	2.18	3 male
t-test	*t*	0.249	0.837	-0.714	
	p	0.772	0.413	0.483	

Table 3.1: Subject characteristics: Descriptive statistics and results of *t*-tests comparing amusic and control participants (N= 11 per group) characteristics. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

All participants were recruited from an existing pool of amusics. Congenital amusia was diagnosed based on the three pitch-based and the rhythm subtest of the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz, Champod, & Hyde, 2003) and a detailed questionnaire about their educational and musical background. Only amusics exhibiting both a pitch perception and a rhythm perception deficit were included in this study to ensure homogeneity as much as possible. MBEA scores as the proportion correct out of 30 and *d'*-scores on all 6 subtest are listed in Table 3.2. Traditionally, the MBEA is scored with proportion correct and amusics are diagnosed based on these scores, but recently criticism has arisen on this scoring method (Henry & McAuley, 2013; Pfeifer & Hamann, 2015). On the one hand, MBEA proportion correct scores are not normally distributed and amusic and control group exhibit different variances. Therefore non-parametric tests would be in order for proportion correct scores, as confirmed by a normality and variances analysis run on our data (but not for *d'*-scores). On the other hand, proportion correct scores have also been shown to lead to misdiagnosis due to participants' response bias (Pfeifer & Hamann, 2015). Therefore, a further analysis using Signal Detection Theory (SDT) (Green & Swets, 1966) was carried out. The SDT measure *d'* is bias free and reflects participants' discriminatory ability without the response bias. The difference between our two groups was calculated based on *d'* values.

Our amusic group falls below traditional proportion correct cut–off scores from Peretz. et al. (2003) on all but the memory subtest, and amusics and controls discriminatory ability is significantly different as shown by the *t*-test values in Table 3.2.

			Scale	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**
Pro-portion Correct	control	Mean	27.45	28.36	28.45	28.55	28.55	28.64
		SD	1.63	1.12	1.29	0.82	1.81	1.69
	amusic	Mean	21.91	20.55	19.82	22.64	21.55	25.18
		SD	2.17	2.42	1.94	3.67	3.96	3.06
d'	control	Mean	3.33	3.61	3.64	3.61	3.76	3.88
		SD	0.76	0.53	0.70	0.53	0.92	0.89
	amusic	Mean	1.41	1.31	1.10	1.80	1.57	2.40
		SD	0.46	0.63	0.54	0.93	1.21	0.94
t-test (df 20)		*t*	-7.13	-9.27	-9.58	-5.63	-4.77	-3.79
		p	.000	.000	.000	.000	.000	.001

Table 3.2 Montreal Battery of Evaluation of Amusia Scores: Means and standard deviations of proportion correct scores (number of correct responses, out of 30), and *d'* scores for amusics and controls (N= 11 per group). The t-test was calculated on *d*'-values. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

All participants were tested in a sound-attenuated chamber in the phonetics laboratory at the University of Düsseldorf. The Ethical Committee of the Medical Faculty at the University of Düsseldorf approved the study protocol and each participant signed an informed consent form before the experiment commenced, and received a small monetary reimbursement for their participation afterwards.

2.2. Stimuli

Our stimuli were isolated synthetic vowels based on auditory properties of natural German front mid vowels. We decided to use mid vowels to avoid periphery effects (Polka & Bohn, 2003), and to utilize vowels that are close to each other in their height and front-back dimension in the vowel space, but that differ in quality and/or quantity. Those considerations left us with the German front mid vowels /ɛ/ (short and more open), /ɛː/ (long and more open), /e/ (short and more closed) and /eː/ (long and more closed). All four are vowels occur in German, however only three of them, /ɛ/, /ɛː/ and /eː/, are native to German and therefore regarded as phonemes, at least adopting a standard view (Wiese, 1996), see the contrast *Betten* /ɛ/ 'beds' – *bäten* /ɛː/ 'if they requested' – *beten* /eː/ 'to pray' (Kohler, 1990), while /e/ occurs only in loanwords in unstressed position (e.g. in *Chemie* 'chemistry'). However, /e/ is also considered to be a phoneme by some authors at least (Giegerich, 1985; Wurzel, 1981). In addition, the occurrence of /ɛː/ is also rather restricted, and many speakers replace this sound by /eː/ (Moulton, 1962). We also considered the phoneme frequency based on Aichert et al. (2005), who in turn based their analysis on CELEX (Baayen, Piepenbrock & Gulikers, 1995). /ɛ/ occupies the 5th rank among vowels and /eː/ the 7th, however, /ɛː/ is ranked 16th (out of 19) and /e/ is not considered as a German phoneme and does therefore not appear at all. Pätzold and Simpson (1997) offer similar frequencies based on the Kiel corpus: /ɛ/ and /eː/ have almost the same frequency, while /e/ and /ɛː/ do not appear in their analysis at all. Regardless of this, we still regarded these four vowels as our best option, as no other German four-way vowel contrasts are as ideally spaced in the vowel triangle, while avoiding periphery effects at the same time.

The durational and formant values for the short, more open vowel /ɛ/ and the long, more closed vowel /eː/ that we employed as basis for the creation of our stimuli can be found in Table 3.3 (based on acoustic measurements by Jessen, 1993).

Vowel	Duration (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/e:/	110	350	2157	2793
/ɛ/	60	524	1869	2624

Table 3.3: Auditory properties of vowels: Duration and formant values of the two German vowels /e:/ and /ɛ/ from Jessen (1993).

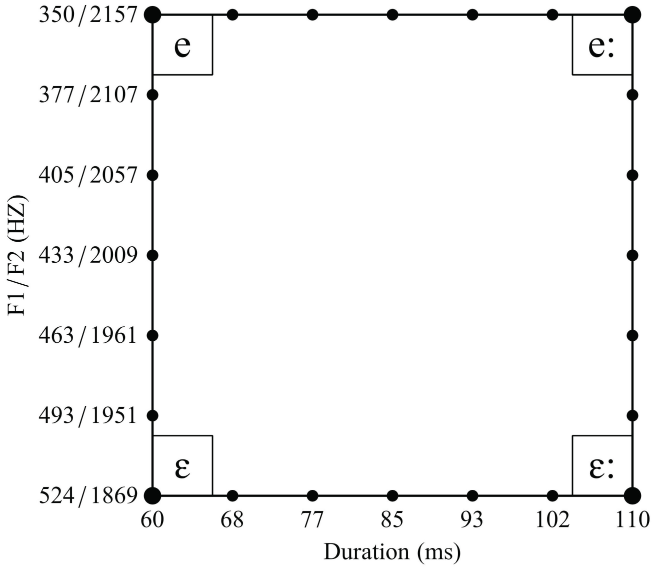


Figure 3.1: Spectral and durational values of stimuli: The four corner vowels with the acoustic properties of real German vowels, and 20 stimuli on four continua in between.

Based on these vowels, we created four continua with seven steps each, varying in either duration or spectral properties. F1 and F2 values were varied together and an F3 value of 2709 Hz, the mean value of the F3 of both vowels, was used for all stimuli. All vowels were created using the Klatt-synthesiser in Praat (Boersma & Weenink, 2016). We used synthesized vowels to ensure very tightly controlled stimuli (Iverson, 2012), while at the same time, we tried to keep them as naturally sounding as possible by adding a falling-rising pitch contour and amplitude, and seven additional formants. The continua and stimuli spacing are based on Escudero & Boersma (2004). The seven spectral values were equally spaced on a mel scale, and the seven durational values were equally varied along a logarithmic scale. The four vowels on the corners of the continua, depicted in Figure 3.1, and the fives steps in-between for each continuum yielded a total of 24 distinct vowel stimuli (4 corner vowels plus 4 x 5mvowels in-between). The four continua and their acoustic properties that were varied are depicted in Figure 3.1. There were two durational continua, one from /e/ to /eː/ and one from /ɛ/ to /ɛː/, and two spectral continua, one from /e/ to /ɛ/ and one from /eː/ to /ɛː/. All 24 vowels were used in the behavioral experiment, while only the four corner vowels with naturally occurring values served as stimuli in the EEG study.

2.3. Behavioral paradigm

An AXB forced-choice discrimination task was used with each continuum. A and B were always two adjacent corner vowels, while X could be either one of the stimuli in between or one of the two corner vowels. Each corner vowel appeared both as A and as B for every continuum, leading to 14 trials for each of the four continua. Each trial was repeated five times, resulting in a block of 280 trials. The order of all trials was pseudo-randomized for participants, so that the same trial was not repeated twice after one another within the block.

In addition, two different inter-stimulus intervals (henceforth: ISI), namely 0.2 s and 1.2 s, were used in order to assess the rapid-auditory processing of amusics (Williamson et al., 2010) and a possible difference between auditory vs. phonemic processing (Werker & Logan, 1985). The ISI was kept constant within the block of 280 trials described above, yielding two blocks. The order of blocks was counterbalanced across participants.

The participants' task was to judge whether the sound in the middle was more similar to the first or to the last sound. They were asked to listen carefully to each trial and to respond as quickly as possible by pressing keyboard buttons denoted as first or last. This procedure was chosen in order to avoid explicit categorization or any influence of orthography.

Participants were seated in a sound-attenuated booth, and the stimuli were presented over AKG K 601 headphones using Praat (Boersma & Weenink, 2016) on a Windows XP computer. Participants could adjust the volume to a comfortable level. The behavioural session took approximately 45 to 60 minutes with one longer break in between, and the possibility to take a short break after every 40 trials.

2.4. EEG Paradigm

The EEG session took place approximately a year after the behavioral session and lasted about 3 hours in total, of which about 1.5 hours were EEG recording time. Each session consisted of four approximately 22-minute recording blocks with short breaks in between. Participants completed a passive listening task while watching a silenced nature documentary without subtitles or visible lip movement, as lip movement from a different language than the one being tested has been shown to interfere with perception (Kang, Johnson, & Finley, 2016; Shinozaki, Hiroe, Sato, Nagamine, & Sekiyama, 2016). Participants were instructed to disregard the sounds they were exposed to and to focus their attention on the movie. The auditory stimuli were presented at 60 dB via two loudspeakers placed in front of the participant at a distance of approximately 1 m.

The auditory stimuli were presented in a multi-deviant oddball paradigm with four blocks, which were counterbalanced across participants. A total of 3600 stimuli occurred in each block. In each block, one corner vowel was the standard, while the other three corner vowels served as deviants. The standard occurred 85 % of the time and each deviant occurred 5 % of the time. Each block started with 20 standards, followed by the oddball sequence, where each deviant was separated from the next by at least four and at most eight standards. The ISI was varied randomly between 400 ms and 600 ms to avoid entrainment effects to the stimulus chain (Repp & Su, 2013; Tal et al., 2017).

Across the four blocks, each vowel occurred once as standard and three times as deviant. The MMN was calculated per participant by subtracting the average event-related potential (ERP) of the standard from each of the three deviants per block, resulting in MMNs for 12 different conditions: The four vowels occurring in three types of contrasts each, namely a formant contrast, a duration contrast and a formant and duration contrast simultaneously. Per condition, the most negative peak in the 100 to 250 ms after stimulus onset was determined. Each MMN amplitude was calculated as the mean voltage over a 40 ms time window centred at the most negative peak.

2.5. EEG recording and pre-processing

The EEG was recorded using a BioSemi Active Two system (BioSemi

Instrumentation BV, Amsterdam, The Netherlands) with 64 Ag-AgCl electrodes that were placed according to the international 10/20-system in a cap fitting the participant's head size. 7 further electrodes were placed on the tip of the nose, the left and right mastoid, below and above the left eye and the outer canthi of the left and right eyes (recording the electro-oculogram (EOG)). The EEG signal was recorded at 8192 Hz and later down-sampled to 512 Hz. The subsequent analyses were performed in Praat (Boersma and Weenink, 2016). The data were offline referenced to the average of the two mastoid channels. Slow drifts were removed by subtracting a line from each channel so that the first and the last sample become zero. The data were bandpass filtered in the frequency domain with a low cut-off of 1 Hz (0.5 Hz bandwidth) and a high cut-off of 25 Hz (12.5 Hz bandwidth). The EEG was segmented into epochs of 500 ms, from 110 ms before to 390 ms after stimulus onset. For baseline correction, the mean voltage of the 110 ms pre-stimulus served as a baseline for amplitude measurement and was subtracted from each sample in this epoch. Artifact correction was done automatically and epochs with an EEG or EOG change exceeding +/−75 µV were excluded. Participants with more than 40% of artifact-contaminated epochs would have been excluded from analysis, though no participant exceeded this limit.

2.6. Scoring and Statistical Analysis

As already mentioned in the participants section, we scored the MBEA using Signal Detection Theory (Green & Swets, 1966), since its measure *d'* reflects participants' discriminatory ability without response bias. We therefore decided to also use *d'* for the scoring of our behavioral data.

In favor of using classical statistics such as ANOVAS, we opted for constructing linear mixed effects regression models using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) written in the statistical computing language R (R core team, 2016) for both our behavioral and our EEG data. Recently, multiple studies have employed mixed models for the analysis of ERP data as well (Bagiella, Sloan, & Heitjan, 2000; Frömer, Maier, & Abdel Rahman, 2018; McWhinney, 2018; Winsler, Midgley, Grainger, & Holcomb, 2018).

So-called mixed models offer the possibility of including fixed effects and random effects. The former are the predictor variables under investigation, which are assumed to be fixed and measured without error

(Bagiella, et al., 2000). The latter represent random samples from a distribution of values, meaning they estimate systematic variance between, for example, individual participants and adding these to a model can help to account for unpredictable variability that arises from these sources and which in turn can lead to biases in group-level estimates (Bagiella, et al., 2000; Frömer, et al., 2018; McWhinney, 2018). This is an important consideration as not all experimental manipulations yield uniform effects across all participants, resulting in individual variation. The benefit of accounting for such individual variation by adding random effects is an improved model accuracy (Bagiella, et al. 2000; McWhinney, 2018). Adding, for example, the variable participant as a random effect allows to resolve the issue of independence among repeated measures by controlling for individual variation among participants (Koerner & Zhang, 2017).

Following established modeling practices (Baayen, Davidson, & Bates, 2008), a full model that included all variables and interactions was used as a starting point. Stimuli were contrast coded and only explicit contrasts were used, i.e. variables were centered around zero. Then it was checked by visual inspection that the residuals of the model were normally distributed and did not reveal heteroscedasticity, i.e. an inconsistency in variability within groups of a categorical variable. The regression models were then simplified by the stepwise exclusion of non-significant fixed effects. A fixed effect was considered non-significant if its *p*-value was higher than 0.05, there were no significant interactions including it and if the Akaike Information Criterion (AIC) of the model including the predictor was higher than when the predictor was not included. A higher AIC shows that a model with this predictor has a smaller explanatory power than a model without the predictor variable in it. Only the final model will be reported in the following sections.

3. Results

3.1. Behavioral

The data were scored by counting the response to the target stimulus in relation to the first stimulus per continuum and then calculating the percentage correct.

We performed several analyses on the data: First the mean per condition per continuum and per ISI was calculated, then we calculated the reliances and weights and found the categorical boundaries per group. We visually inspected the data and performed a linear mixed model.

The mean discrimination curve averaged across all conditions is depicted in Figure 3.2. In general, controls exhibited a steeper categorization curve than amusics. Figure 3.3 depicts the categorization curves split per continuum and inter-stimulus interval. Here, a steep slope in the category boundary for spectral cues (two panels on the right) shows good performance of both amusics and controls. The slope for the discrimination of durational cues (two panels on the left) is not as steep, especially for the short ISI.

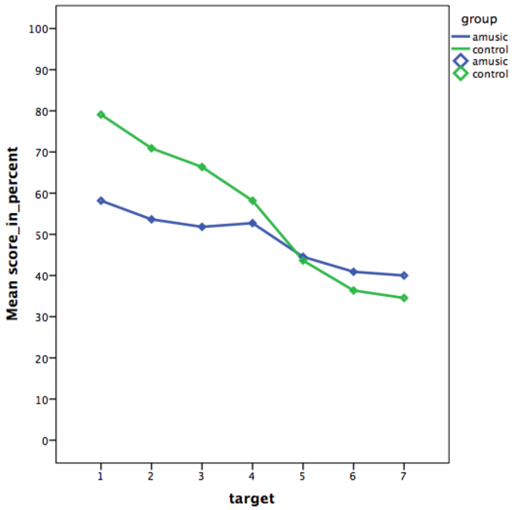


Figure 3.2: Discrimination curve of amusics (blue/darker) compared to controls (green/lighter) averaged across all conditions.

Next, we computed the individual reliances and weights per participant and per group, following Escudero et al.'s (2004) procedure: For each reliance we computed the percentage of correct answers at one edge of the continuum minus the percentage at the other edge.

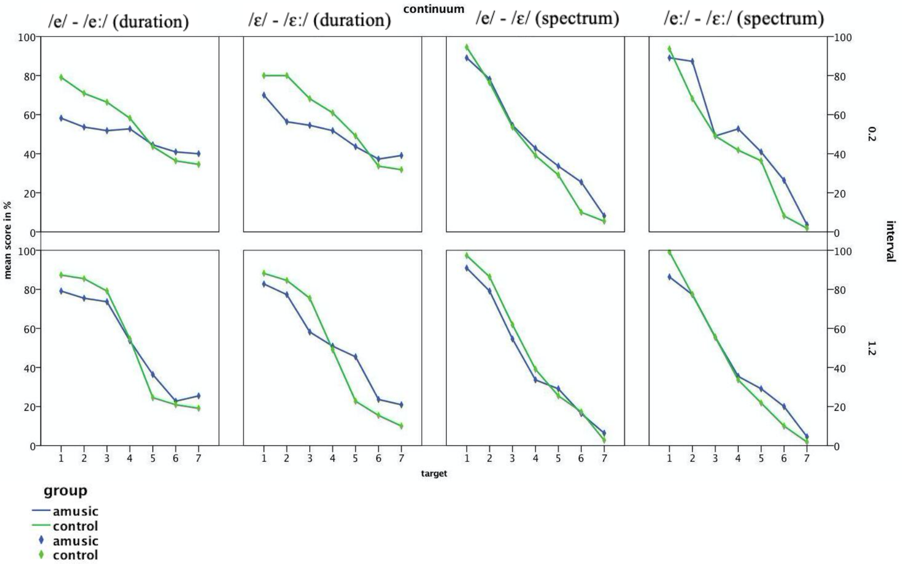


Figure 3.3: Discrimination curves of controls compared to amusics split by long and short inter-stimulus interval and by condition.

The cue weighting yielded the following results: 7 of 11 controls relied somewhat more on spectral cues (between 20 – 32 % difference); 3 relied mostly (40 % and more difference) on spectral cues and only 1 relied equally strong on both cues (within 10%). For the amusics, a slightly different picture emerged: 8 out of 11 relied mostly on spectral cues (40 % and more difference); 1 relied more on spectral than on durational cues and another 2 relied equally strong on both (within 10%), cf. Table 3.4.

As a next step, we computed *d’* and calculated a mixed model based on all trials and a second one based on edges only. The reasoning for this was that across all trials, some more ambiguous stimuli were also included making the task more difficult for amusics and controls alike. In a second analysis with only the edges included, i.e. the most natural and unambiguous stimuli, we expected controls to perform almost at ceiling, while it was the question whether amusics would still exhibit significant deficits.

Subject	Duration reliance in %	Spectrum reliance in %	Group
A10	0	78	Mostly spectrum; duration < 50%
A5	8	60	
C9	13	83	
A8	23	88	
C3	23	98	
A9	30	88	
A2	38	88	
A11	38	85	
A3	40	90	
C8	40	90	
A1	48	90	
C5	60	80	Both but more spectrum
C6	63	93	
C2	68	100	
C4	68	95	
A7	73	93	
C1	78	98	
C11	78	98	
C10	80	98	
A4	85	95	Equal within 10 %
A6	73	63	
C7	90	95	

Table 3.4: Durational and spectral reliances (cells of amusic participants shaded)

A linear mixed model with subjects as random effects and everything else as fixed effects (i.e. the complete model) gave the following results: Amusics performed worse than controls ($t(20) = 2.28$, $p = 0.033$). Durational cues were overall harder to perceive than spectral cues ($t(1028) = 8.24$, $p < 0.001$), and shorter ISI was overall harder for both groups ($t(1028) = 7.69$, $p < 0.001$). We also found two interactions: Group by cue ($t(1028) = -2.18$ $p = 0.029$), with amusic performing worse for duration than controls, and cue by interval ($t(1028) = -4.689$, $p < 0.001$), with duration being harder with a short ISI.

A model with the same parameters across edges revealed almost the same results: Amusics performed worse than controls ($t(20) = 2.14$, $p = 0.0045$), duration was overall harder than spectral cues ($t(324) = 14.44$, $p < 0.001$) and a shorter ISI was overall harder ($t(324) = 5.86$, $p < 0.001$). The group by cue interaction was no longer significant. The other interaction, cue by interval, remained significant ($t(1324) = -5.02$, $p < 0.001$). Figure 3.4 depicts the main effect of group across all trials (left panel) and for the edges only (right panel) and Figure 3.5 shows the absence of the interaction between group and cue (left panel) and the presence of an interaction between cue and interval, both when considering only the edges.

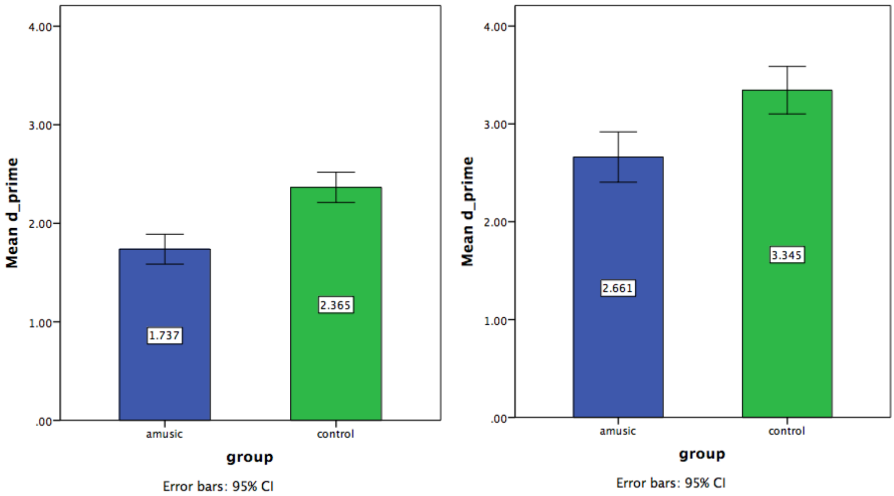


Figure 3.4: Significant main effect of group across all trials (left panel) and for the edges only (right panel)

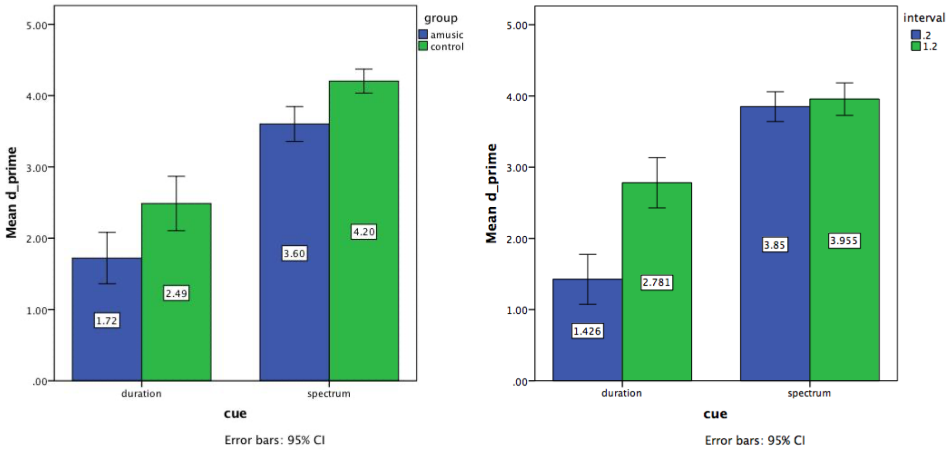


Figure 3.5: Interaction in regression across edges. Left: No interaction between group and cue. Right: Significant interaction between cue and interval.

3.2. EEG

This analysis was run on the MMN amplitude measured at 9 channels (Fz, FCz, Cz, F3, F4, FC3, FC4, C3, C4).

Visual analysis confirms the negative polarity, the expected latency and fronto-central scalp distribution of the MMN. Figure 3.6 shows the average difference waveform averaged across all conditions for amusics (dotted line) and controls (solid line) and the scalp topography (the darker blue, the more negative). The figure also confirms a difference between amusics and controls.

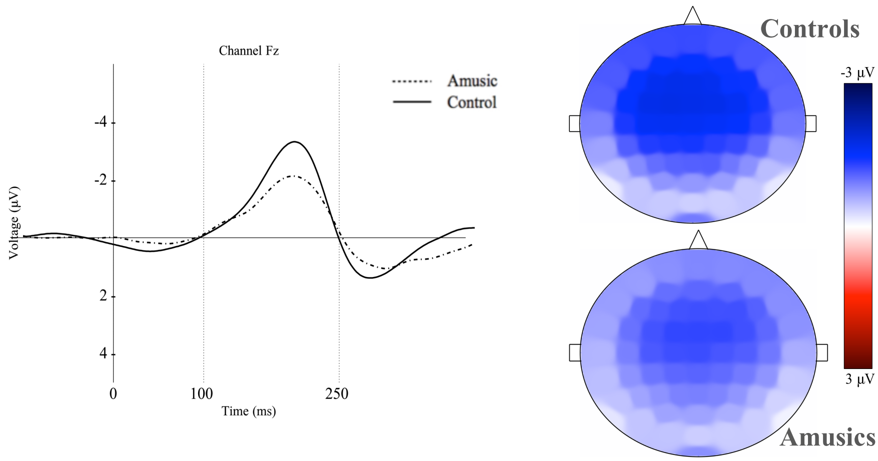


Figure 3.6: Difference curves (deviant - standard) for amusics and controls in a time window between 100 and 250 ms averaged across all conditions. On the left are the grand average difference waves plotted at Fz, and on the right topographical maps.

Figure 3.7 shows the grand average difference waves between 100 and 250 ms plotted at Fz for amusics and controls split per condition, and Figure 3.8 shows the corresponding scalp topography in the same time window. Both figures illustrate the, on average, greater negativity in the control group, depicted in Figure 3.7 by the solid line for the control group and in Figure 3.8 by the darker blue upper scalp topography per cell. The scalp topography also reveals a slightly less specified distribution in the amusics. Both figures also seem to indicate differences between the conditions, as seen for example in the much lighter colored scalp topographies for the standards /e:/ and /ɛ:/ with durational deviants in Figure 3.8.

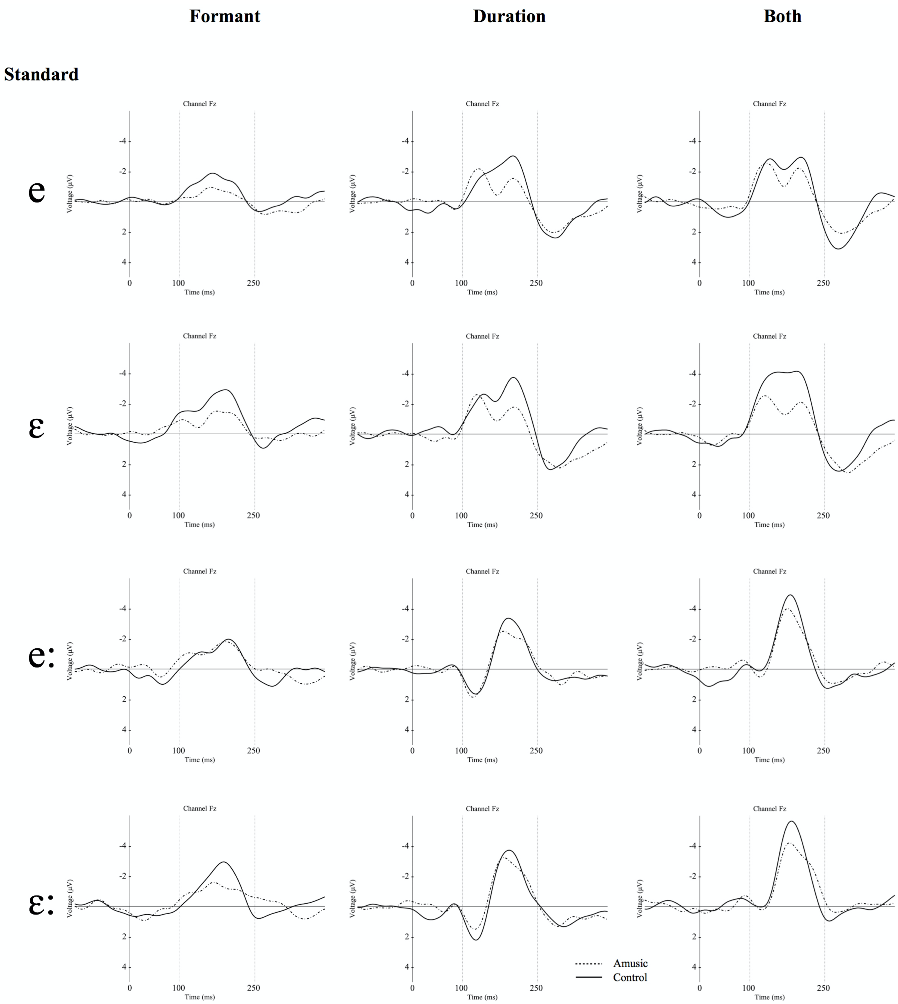


Figure 3.7: Grand average difference waves plotted at Fz for amusics and controls per condition.

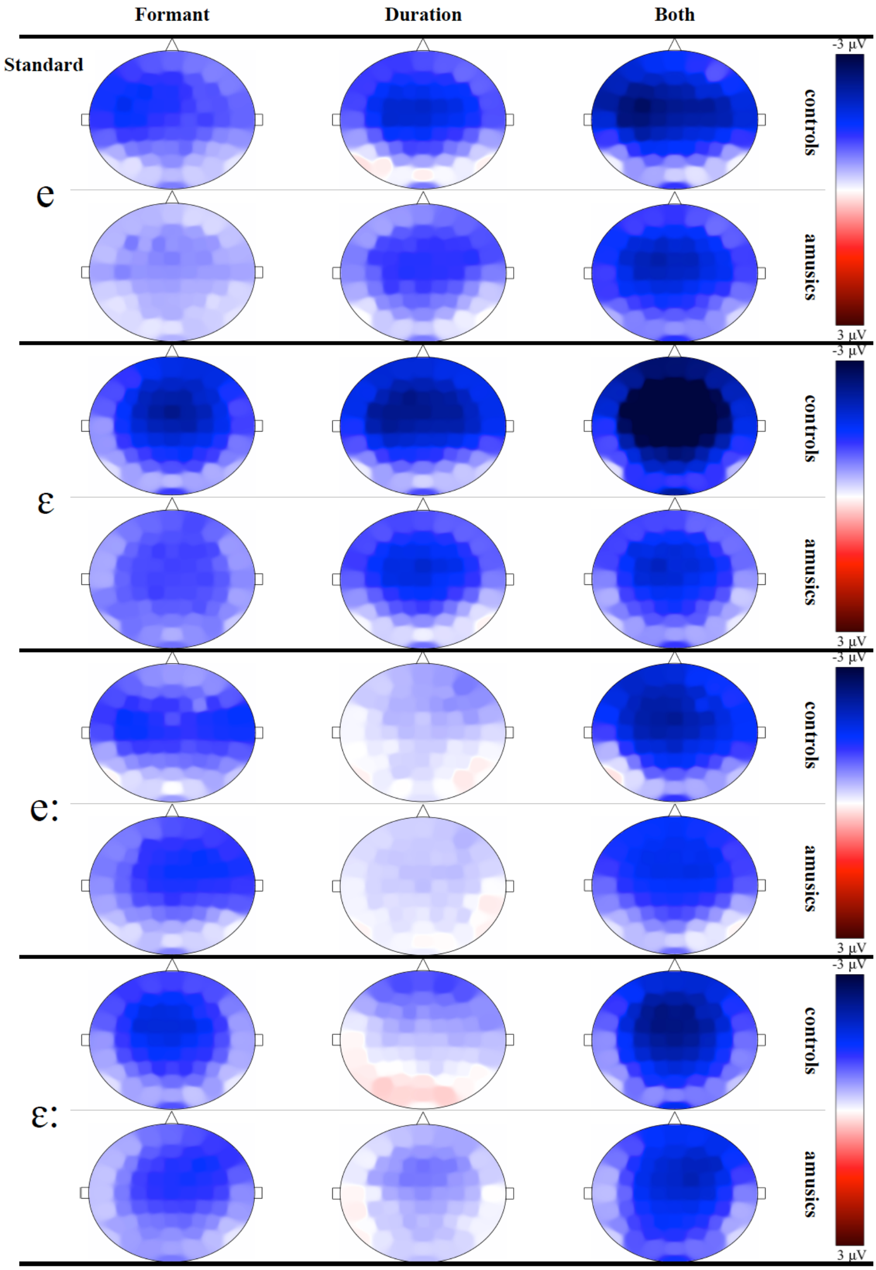


Figure 3.8: Topographical maps of amusics and controls averaged in a time window between 100 and 250 ms per condition, with controls at the top and amusics at the bottom.

After visual analysis, we performed two-tailed *t*-tests against zero separately for each group to determine whether the difference waveform response was present in every condition, all of which were significant. Significance levels and mean values at Fz are given in Table 3.5.

Standard	Formant	Duration	Both
/e/	-2.15*** -1.40***	-3.15*** -2.34***	-3.86*** -3.09***
/ɛː/	-2.98*** -1.96**	-3.84*** -2.69***	-4.63*** -2.94***
/eː/	-2.18*** -2.38***	-2.83*** -2.40***	-4.03*** -3.75***
/ɛː/	-2.60*** -2.17*	-3.22*** -2.93***	-4.71*** -3.98***

Table 3.5: Mean voltage at electrode Fz measured in μV. Top value is that of the control group, bottom value that of the amusic group. * indicate significance levels in t-tests against zero: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Next, we performed an independent samples *t*-test between the groups, which was also highly significant ($t(2374) = -9.59$, $p < 0.001$).

As we were interested in the differences between amusics and controls in every condition, we calculated a linear mixed model (lmer) in R with subject as random effects and group and condition as fixed factors. We found significant main effects for group ($t(23.7) = -2.43$, $p = 0.023$) with amusics (M = -2.67) overall having a smaller MMN than controls (M= -3.35). A visualisation of this is given in Figure 3.6 above, depicting the grand average difference waves plotted at Fz and topographical maps of amusics and controls, which shows a broadly distributed MMN. In addition, we found a main effect for condition ($t(2351.8) = -6.14$, $p < 0.001$) and a significant interaction between group and condition ($t(2351.8) = 3.85$, $p < 0.001$). To understand the Group by Condition interaction, separate t-tests were performed for each condition separately. They all yielded significant results, i.e. controls having a bigger MMN, in all but one condition, from standard /eː/ to deviant /ɛː/. A visualisation can be found in the topographical plots in Figure 3.8 above.

4. Discussion

Our studies showed that amusics perform behaviorally worse than controls in the discrimination of German vowels based on spectral and durational cues. And despite the fact that they exhibit a neural response of automatic change detection, this response is significantly reduced in comparison to controls.

The behavioral findings are in line with our expectations that amusics should show impairments in spectral and durational cue perception, as only participants' with a deficit both in pitch and rhythm perception were included in this study.

When ambiguous stimuli were included, amusics struggled more with durational cues than with spectral cues in comparison to controls, as shown by a significant interaction. When only stimuli with the original, natural formant frequencies were included, this interaction vanished. However amusics still performed significantly worse even for those most natural vowel stimuli. Both amusics and controls performed worse at a short inter-stimulus interval. For the amusics, no special deficits in rapid-auditory processing were observed that were expected based on previous findings by Williamson et al. (2010).

Concerning the inter-stimulus interval, another interaction was observable: Stimuli with a durational difference were harder to discriminate than those with a spectral difference in the short inter-stimulus interval condition. For spectral differences, the inter-stimulus interval did not make a difference in the discrimination. There are two possible explanations for this interaction: The durational difference between /e/ and /e:/ might not have been as salient as the spectral difference between /e/ and /ɛ/ in our stimuli, and therefore might have been harder to perceive, especially with a shorter ISI. Another explanation might be that in order for the phonemic processing to occur and to enable a differentiation between vowel phonemes, a longer ISI is needed, as described by Werker and Logan (1985).

The additionally calculated discrimination curves support the picture described above: Amusics and controls display steeper discrimination curves, indicating better discrimination, for spectral than for durational cues. Especially shallow curves emerged for durational cues at the short inter-stimulus interval.

When considering the reliance scores, however, a slightly different picture emerged: The majority (6 out of 11) of control participants relied equally strong on durational and spectral cues, while the other controls relied more or mostly on spectral cues. The majority of the amusics (7 of 11), on the other hand, relied most or mostly on spectral cues, only 3 relied equally strong on both cues and 1 relied more on duration.

These are all new findings as none of the previous studies on amusics and vowel perception have investigated durational cues at all. All were conducted with native speakers of a tonal language and mostly considered linguistic tone at the same time. In addition, Huang et al. (2015) found that at least the speech tone deficits in Mandarin amusics is independent of duration, however there is no evidence that the same holds true for vowel differences. Despite this, Tang et al.'s (2018) vowel stimuli ranged between 236 ms and 772 ms, which could have served as perceptual cue for amusics, explaining why their performance was so comparable to that of the controls.

The findings of Huang et al. (2016) that amusics exhibit an overall lower discrimination rate in comparison to controls can be confirmed by our data. Zhang et al.'s (2017) conclusion that the deficit in amusia is a general processing deficit of spectral frequencies can only partly be supported. While the amusics in our study did indeed also have a processing deficit for spectral cues, they still relied on spectral cues more heavily than on durational cues if the reliance rates of amusics and controls are compared. And lastly, all three studies employed only behavioral tasks, while our findings are also supported by the electrophysiology underlying such behavior.

Our ERP results show that the auditory system of amusics has formed a representation of the repetitive aspect of the standard stimulus as represented by the MMN that they exhibit. However, their MMN is significantly reduced in comparison to our control population and might therefore represent a more inaccurate discrimination of the stimuli than the controls' larger MMN does (Kujala & Näätänen, 2001). The general finding that amusics did indeed display an MMN response is in direct opposition to that by Braun et al. (2008), who did not find an MMN in amusics at all. Furthermore, the fact that the response of our amusics was significantly reduced is in opposition to the findings by Moreau et al. (2009, 2013) and Mignault et al. (2012) who found amusics displayed completely normal MMNs. All these findings utilized musical stimuli, however. Our study supports Nan et al.'s (2016) results, who also found a reduced MMN for tone-language speaking amusics in response to lexical tones. Our findings are also in line with Partanen et al. (2011) and Ylinen et al. (2005), as durational changes elicited bigger MMN amplitudes than vowel quality changes, and simultaneous changes in quantity and quality elicited the biggest amplitude.

In addition, the results of the laterality analysis show a more left lateralized MMN for controls, which is to be expected for a linguistic MMN (Shtyrov et al., 2000; Sorokin, et al., 2010), whereas amusics exhibit a more right-lateralized MMN, which has been reported for different disorders such as dyslexia (e.g. Cutini et al. 2016).

A point of criticism concerning our event-related potential analysis could be the fact that we calculated our MMN by subtracting the average ERP of the standard from each of the three deviants per block, which can be seen as an auditory MMN. Another option would have been to calculate the MMN by subtracting the average ERP of the standard from the three deviant ERPs of the same vowel in different blocks. This would have ensured the resulting difference wave reflected the phonological contrast rather than the acoustic difference between vowels within a block. When planning this study, however, we only considered the analysis of the auditory MMN and we refrained from re-analyzing our data afterwards.

Another criticism concerning our methodology might be the small sample size for an ERP study. This is a pitfall that all studies with amusics face, as the population is a rather small one. However, when considering other ERP studies with amusics, our sample size is fairly large in comparison.

Our study could also be criticized twofold concerning the choice of stimuli: Firstly, we could have chosen natural recordings of vowels and manipulated those instead of synthesising ones to make absolutely sure our vowels were perceived as such and not purely auditory as non-speech sounds. However, we opted for synthesized vowels as we favored to have control over all parameters, while we ensured at the same time that the vowels sounded as natural as possible (see methods section for details and Iverson 2012 for a discussion about this topic). Secondly, we could have chosen four vowels that are considered as German native phonemes by traditional accounts of German phonology such as Wiese (1996), and where all four exhibit similar phoneme frequencies. The choice of the presently employed mid vowels, however, avoided periphery effects, and two of them, /e:/ and /ɛ/, at least offer a very similar, and relatively high, frequency of occurrence that other vowel pairs might not have offered.

We also chose vowels in isolation on purpose, even though one might argue that syllables with transitional cues or whole words with meaning might have been better suited. We wanted to focus purely on spectral and durational vowels cues, while disregarding other cues for the moment, hence our choice.

5. Conclusion

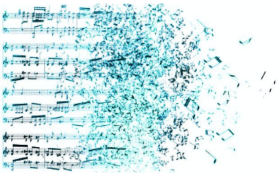
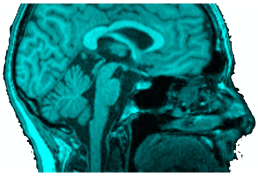
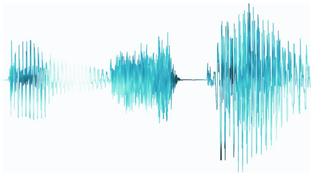
Our study showed for the first time that amusics perform behaviorally worse in discriminating vowel pairs based on spectral as well as durational differences. We could not find evidence to support a rapid-auditory processing deficit in amusics, but rather found that a shorter inter-stimulus interval made processing harder for amusics and controls alike.

We were also able to demonstrate that amusics did indeed show an MMN, as an automatic reaction to auditory change detection to different vowels. However, this MMN was significantly reduced in comparison to our control population, indicating abnormal neural processes even at this very early stage of processing.

Taken together, these findings show that amusics exhibit difficulties when it comes to vowel processing at least in isolation. Further studies investigating later ERP components, as well as more complex linguistic stimuli such as syllables or words are therefore warranted.

6. Acknowledgements

We thank Dirk Vet for helping us writing the E-Prime script to test our participants.

4. Word Stress Perception by Congenital Amusics

<u>This chapter is based on:</u>

Pfeifer, Jasmin & Silke Hamann. Word Stress Perception by Congenital Amusics
(accepted as a chapter in: *How Language Speaks to Music,* eds: Mathias Scharinger, Richard Wiese; Mouton de Gruyter.)

Abstract[1]

Congenital Amusia is a developmental disorder that is defined by difficulties with the perception of pitch and rhythm. While it used to be described as a disorder of musical pitch perception, recent publications have shown that congenital amusia also affects linguistic pitch perception. In this chapter we report one of the first study of word stress processing by congenital amusics. We designed a behavioral identification task and a mismatch negativity study using German minimal stress pairs as basis for our stimuli. We considered the acoustic parameters pitch, duration, intensity and spectral slope. Behavioral results surprisingly revealed no pitch processing difficulties in the amusic group and showed a better usage of durational cues in the amusic group. The electrophysiological results revealed that amusics consistently have an MMN, though it is significantly smaller than that of controls. The present results warrant further investigation of the use of linguistic cues by congenital amusics.

1. Introduction

1.1. What is congenital amusia?

Congenital amusia (henceforth: Amusia) is a neuro-developmental disorder that has a negative influence on pitch perception and partly also on rhythm perception (Peretz et al., 2002; Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Stewart, 2008). Amusia is neither caused by insufficient exposure to music, nor by a hearing deficiency, brain damage or intellectual impairment (Ayotte et al., 2002). People with amusia (in the following called amusics) face lifelong impairments in the musical domain. Their symptoms can range from an inability to discriminate notes of different pitches, an inability to recognize well-known songs without lyrics or an inability to recognize out of tune singing, to an inability to recognize music as such. In the most extreme cases, their symptoms can be so severe that music causes discomfort to them (Peretz et al., 2002; Foxton et al., 2004; Stewart, 2008). Most likely due to those more

[1] Parts of the abstract and the introduction of this chapter are identical to parts of the introduction sections in Hamann et al. (2012) and Pfeifer et al. (2014) (this holds for the short definition of amusia and its symptoms, the description of possible causes of amusia and of the MBEA, and the summaries of the studies by Patel et al. 2008 and Liu et al. 2010). These sections in the original studies were written by Jasmin Pfeifer and Silke Hamann.

apparent symptoms, early research has mostly been focused on the influence of amusia on music. Hence, amusia has long been characterized as a music-specific disorder (Ayotte et al., 2002; Peretz et al., 2002; Peretz, Blood, Penhune, & Zarorre, 2001). Different aspects of musical engagement have been assessed and found impaired in amusia, such as pitch perception (Peretz et al., 2002), pitch production (Dalla Bella, Berkowska, & Sowiński, 2011), rhythm perception (Foxton, Nandy, & Griffiths, 2006), beat synchronization (Sowiński & Dalla Bella, 2013), timbre perception (Marin, Gingras, & Stewart, 2012), consonance rating (Ayotte et al., 2002), and musical emotion perception (Marin, Thompson, & Stewart, 2012).

Amusia is neither caused by insufficient exposure to music, nor by a hearing deficiency, brain damage or intellectual impairment (Ayotte et al., 2002), and the disorder is considered innate (e.g. Peretz et al., 2002; Ayotte et al., 2002). The underlying cause of this multi-faceted disorder has been hypothesized to be a fine-grained pitch processing deficit (Ayotte et al., 2002; Foxton et al., 2004; Hutchins, Gosselin, & Peretz, 2010; Hyde & Peretz, 2004), a pitch memory deficit (Gosselin, Jolicœur, & Peretz, 2009; Tillmann, Schulze & Foxton, 2009; Williamson & Stewart, 2010; Tillmann, Lalitte, Albouy, Caclin & Bigand, 2016), a statistical learning deficit (Peretz, Saffran, Schön, & Gosselin, 2012) or a rapid-auditory processing deficit (Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010, Albouy, Cousineau, Caclin, Tillmann & Peretz, 2016) and partly also a rhythm/beat perception deficit (Phillips-Silver, Toiviainen, Gosselin, & Peretz, 2013; Launay, Grube, & Stewart, 2014). However, there is no consensus yet on the cause, and it is likely a multi-causal deficit that is responsible for the different symptoms exhibited by amusics. The exact neural underpinnings are also still unknown and various studies implicated different brain areas as having structural or functional abnormalities: Less white matter in the left and right inferior frontal gyrus (IFG; Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006); more grey matter in the right inferior frontal gyrus and right superior temporal gyrus (STG) and other brain areas (Hyde, Lerch, Zatorre, Griffiths, & Evans, 2007); less grey matter in the left IFG (Broca's area) and left STG (Wernicke's area; Mandell, Schulze, & Schlaug, 2007); less white matter in the arcuate fasciculus (AF; a fiber bundle connecting IFG and STG; Loui & Schlaug, 2009); abnormal deactivation and reduced connectivity in the right IFG (Hyde, Zatorre, & Peretz, 2011); more grey matter and less white matter in the right IFG and less grey matter in the right STG (Albouy, et al., 2013b);

decreased magnetic N100 amplitude in the left and right STG and the left and right IFG, and decreased activation in the right dorsolateral prefrontal cortex (DLPFC) for memory tasks (Albouy, et al., 2013a); and abnormal white matter structural connectivity in the right AF (Chen & Yuan, 2016). However, Chen et al. (2014) showed that the findings concerning the detection of the AF might be questionable as they strongly depended on the tracking algorithm that was used.

Due to these mixed and not yet fully substantiated findings, no neurological markers can be used to diagnose amusia. Instead, the behavioral markers mentioned above are currently used to screen for amusics. The most widely used tool for amusia screening is the *Montreal Battery of Evaluation of Amusia* (MBEA; Peretz et al., 2003), a series of tests that was originally devised to assess the musical abilities of brain-damaged patients. Nowadays, it is the main tool used to screen for and diagnose congenital amusia, in combination with questionnaires. It consists of six subtests, namely a scale, contour, interval, rhythm, meter, and memory test. A repeated score of 22 or below out of 30 (22 corresponding to two standard deviations below the mean scores of Peretz et al.'s normal participant group) on at least two of the first four subtests, and a self-reported history of problems with music perception used to be utilized to diagnose amusia (e.g. Foxton et al., 2004; Peretz et al., 2003; Tillmann et al., 2009). However, the cut-off scores, scoring procedures and the use of parametric statistics has recently been criticized (Wise, 2009; Henry & McAuley, 2010, 2013; Pfeifer & Hamann, 2015) which has led to the use of Signal Detection Theory (SDT; Green & Swets, 1966), see section 2.2 below.

1.2. Congenital amusia and speech perception

While early research on amusia focused on musical impairments, more recent work also investigated the impact of amusia on language perception, since pitch also plays an important role in speech. In intonation, pitch is, for example, used to disambiguate questions from statements or to mark focus; while on the word level, pitch (among other things) is used to distinguish words with similar segmental structure but different stress pattern (e.g. English *present* vs. *present*, where underscore denotes the stressed syllable) or to distinguish words with identical segments but different tones (e.g. in tone languages such as Mandarin Chinese). Other linguistic information such as conveying emotion or irony also makes use of pitch. Some of these areas of speech perception have been shown to be affected by amusia, i.e. intonation perception (Patel et al., 2008; Liu et al., 2010; Hamann et al., 2012), tone language perception (Liu, Jiang, Wang, Xu, & Patel, 2015; Liu, Maggu, Lau, & Wong, 2015; Liu, Xu, Patel, Francart, & Jiang, 2012; Tillmann et al., 2011) emotional prosody in language (Lolli, Lewenstein, Basurto, Winnik, & Loui, 2015; Thompson, Marin, & Stewart, 2012) and vowel perception in tonal languages (Huang et al., 2016; Zhang et al., 2017; Tang et al., 2018). Due to these findings, amusia has more recently been described as a domain-general disorder affecting pitch processing in general (Hamann et al., 2012; Liu et al., 2010; Zhang, Shao, & Huang, 2017).

The speech perception studies with most relevance to the present study are one the one hand studies about intonation and on the other studies about the usage of durational cues by amusics. The first to systematically research intonation perception impairments were Patel et al. (2008), who investigated the pitch perception of British English and Canadian French amusics in an AX discrimination task. They utilized cross-spliced statement-question pairs that were further edited to acoustically differ only in the final region of the intonation contour. In addition, Patel et al. used tonal analogs of the statement-question pairs. They found that 30% of the amusics had difficulties discriminating statements from questions based on intonation, while they were able to discriminate the tone analogs based on these sentences well (Patel et al., 2008). These findings were in contrast to all previous studies (such as Ayotte et al., 2002) claiming that linguistic pitch perception was unaffected by amusia.

Liu et al. (2010) investigated the pitch processing of British English amusics in an AX discrimination task using statement-question pairs, nonsense speech analogs and tone analogs. As in the study by Patel et al. (2008), the stimuli retained the final pitch of naturally produced statements or questions. The amusics performed significantly worse than controls on all three stimuli types. Furthermore, amusics performed significantly better on gliding tones than on natural speech, while their discrimination of nonsense-speech was worst, thus showing that amusics have an impaired intonation perception. This result differs from Patel et al.'s (2008) insofar as Liu et al. found an impairment of intonation perception for a subgroup of amusics only.

Hamann et al. (2012) investigated the intonation perception of German amusics in an AX discrimination task. They looked at pitch processing as well, including two tonal analogs. However, they were the first to also consider other linguistic factors such as the length of the stimuli, and the continuity of the pitch curve. Their stimuli consisted of short (3–6 syllables) and long (7–10 syllables) statement-question pairs and were also varied concerning the segmental material. Sentences either consisted only of voiced segments (resulting in a continuous pitch, resulting in a discontinuous intonation contour). Amusics were again shown to be impaired in the discrimination of speech as well as non-speech material. It was also found that amusics as well as controls performed worse for continuous intonation contours; however, the length of the stimuli was not found to have an influence.

The speech perception studies with amusics up to now have almost exclusively focused on intonation and on pitch as a perceptual cue, probably due to the fact that most hypotheses on the underlying deficit of amusia are based on some form of pitch perception deficit. However, there are two recent studies that also investigate the perception and usage of durational cues by amusics. One study considering durational differences was conducted with native speakers of a tonal language and considered vowel differences and linguistic tone at the same time. Huang et al. (2015) found that at least the speech tone deficit in Mandarin amusics is independent of duration. However, there is no evidence that the same holds true for the durational cue in word stress perception. The other study (Jasmin, Dick, Holt & Tierney, 2019) investigated the cue weighting of durational versus pitch cues of amusics. On the one hand they investigated what they called “phonetic cue weighting” using the English minimal pairs *beer* and *pier*, for which they identified voice onset time (VOT) as the durational and primary cue and fundamental frequency of the following vowel as pitch and secondary cue. In addition, they also used a “prosodic cue weighting” paradigm in which two phrases with different stress patterns were used. In this paradigm, pitch was regarded as primary cue and durational lengthening as secondary cue. Jasmin et al. (2019) found no differences between amusics and controls in the phonetic cue weighting. In the prosodic cue weighting however, amusics placed greater emphasis on durational cues than on pitch cues.

Contrastive word stress is an area that has not yet been explored in amusia. However, it seems ideally suited to explore the different speech cues and participants’ sensitivity to them. Weber et al. (2004) showed, for example, that German infants were sensitive to the predominant strong-weak stress pattern of their native language at an age as young as 5 months. In addition, the perception (and production) of word stress and its different perceptual correlates have been assessed and found impaired in many other populations such as children at risk of dyslexia (e.g. De Bree, Wijnen & Zonneveld, 2006; Leong, Hämäläinen, Soltész & Goswami, 2012; Goswami et al., 2013), (children at risk of) SLI (e.g. Gallon, Harris & Van der Lely, 2007; Haake, Kob, Willmes & Domahs, 2013; Fikkert & Penner, 1998) and Down Syndrom (e.g. Pettinato & Verhoeven, 2009). In addition to these behavioral findings, electrophysiological evidence concerning word stress perception in clinical populations is also available, as discussed in the following section.

1.3. The electrophysiology of Congenital Amusia

The mismatch negativity (MMN), an early event-related potential (ERP) component, is especially useful for studies on the electrophysiology of word stress perception. It reflects the neural responses of automatic change detection and its recording does not require attentive action from the participant (Näätänen, Gaillard, & Mäntysalo, 1978), for reviews, see Näätänen (2001); Näätänen, Paavilainen, Rinne, & Alho (2007); Picton, Alain, Otten, Ritter, & Achim (2000). The MMN is generated as the brain's automatic, unconscious response to auditory changes but it also indexes behavioral accuracy e.g. Näätänen et al. (2007). Large MMN amplitudes are elicited by accurate stimulus discrimination, and small MMN amplitudes have been shown to be associated with inaccurate discrimination (Kujala & Näätänen, 2001) in various healthy groups and patients. It peaks at around 100 to 250 ms if the auditory system has formed a representation of the repetitive aspect of the standard stimulus and then a deviant occurs (Näätänen, 2001).

MMN paradigms have widely been used in general auditory but also speech perception research (e.g. Chládková, Escudero, & Lipski, 2013; Kirmse et al., 2008; Näätänen, 2001; Partanen, Vainio, Kujala, & Huotilainen, 2011; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). The linguistic MMN is hypothesized to arise not only from auditory change detection but also from the representation of speech sounds in long term memory that facilitate the discrimination process (e.g. Näätänen, 2001; Partanen et al., 2011), and it has been shown to be more left lateralized (Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000; Sorokin, Alku, & Kujala, 2010). Partanen et al. (2011) have shown that the MMN in linguistic research can be used to establish an auditory discrimination profile, taking into account duration, intensity, pitch and vowel differences. A reduced MMN amplitude is thought to reflect poorer representations of the phonetic categories, which is hypothesized to possibly result from poor language-specific learning of relevant phonetic cues (e.g. Näätänen et al., 2014). Weber et al. (2005) investigated ERP responses of 5-month-old German infants at risk of SLI, finding a significantly reduced MMN to changes in stress patterns. These were interpreted as indicating a less effective processing of word stress and thereby leading to a delay in language acquisition. A reduced mismatch negativity was also found in schizophrenia, when investigating stimuli that deviated in frequency, duration and intensity (Hay et al., 2015).

The MMN has been shown to originate in the auditory cortex and the fronto-central scalp areas (for a review see Näätänen et al., 2007), and in a lesion study (Alain, Woods, & Knight, 1998) specifically the dorsolateral prefrontal cortices were implicated. These are also the regions that are affected by amusia, as shown by Albouy et al. (2013b) and Hyde, Zatorre, & Peretz (2011). The MMN therefore seems well suited to investigate the neurophysiological processes underlying congenital amusia.

Numerous studies on pitch perception by amusics have utilized the MMN already, however with widely differing findings: The first to use it were Braun et al. (2008), who found the MMN to be absent in amusics (for melodies containing altered notes). Studies by Moreau et al. (2009), testing adults with melodies, and Mignault et al. (2012), testing children with tonal sequences, on the other hand, found normal MMNs in amusics. The same holds for the study by Moreau et al. (2013), testing amusics with piano tone sequences. Reduced, abnormal MMNs were found by Lebrun et al. (2012) (for one child to small tonal changes of 25 per cent) and by Nan et al. (2016) (for the responses of tone-language speaking amusics to lexical tones). The latter found a reduced MMN only for a subgroup of amusics, namely for tone agnosics. Taken together, the aforementioned findings seem to show absent or reduced MMNs to tonal sequences in at least some amusics.

Based on this, the present study looks at a fairly homogeneous group of amusics (showing both a pitch and a rhythm deficit) of a non-tonal language, German, with a contrastive stress difference. We hypothesize that these amusics will perform behaviorally worse in identifying carefully controlled stimuli based on stress minimal pairs and will show reduced MMNs in comparison to controls. To test these hypotheses, we designed a behavioral study and an electrophysiological one. Before we present the details of this study, we briefly describe the perceptual cues to contrastive word stress in German.

1.4. Perceptual cues to word stress in German

The primary perceptual cue for word stress in German is the duration of the vowel or syllable, with stressed vowels and syllables having longer duration than their unstressed counterparts (Dogil 1995; Jessen, Marasek, Schneider & Classen, 1995; Haake et al., 2013). Secondary cues are pitch, intensity and spectral slope (Lintfert, 2010): Stressed syllables are usually produced with higher pitch than unstressed syllables (Dogil, 1995; Jessen et al., 1995) and they usually are louder (have a higher intensity) than unstressed syllables; however, this strongly correlates with the slope of the frequency spectrum: In stressed syllables, higher frequencies are also produced with a higher intensity, resulting in a less tilted spectrum than in unstressed syllables (Jessen et al., 1995). As unstressed vowels are usually also more reduced than stressed vowels, even in a language like German, where vowel reduction is minimal, the formant values of unstressed values are more centralized than that of stressed vowels (Lintfert, 2010). As mentioned by Haake et al. (2013), all of these cues are highly variable within and across speakers.

2. Materials and Methods

2.1. Stimuli

For the creation of the stimuli, we recorded a native speaker of German producing several repetitions of the stress minimal pair *<u>um</u>stellen* [ˈʊmʃtɛln̩] "to reposition" vs. *um<u>stel</u>len* "to surround" [ʊmˈʃtɛln̩] (where underscore denotes the syllable with main stress), first within a sentence and then in isolation. We then picked the isolated productions that were pronounced the clearest.

For the acoustic analysis and manipulation, we considered each word as consisting of two parts: The first being the verbal prefix, i.e. the first syllable [ʊm] (which in the unstressed cases was mostly realized as a nasalized vowel [ʊ̃]), and the second part being the stem, i.e. the second and third syllable together, excluding the initial fricative, thus [tɛln̩]. For each part of each word, we measured the acoustic parameters duration, pitch, spectral slope and intensity, as given in Table 4.1. Spectral slope (or tilt) was calculated as the slope between the low frequency band (below 1 kHz) and the high frequency band (from 1 to 4 kHz). The realization of the fricative [ʃ] did not differ in any of the parameters between the two words and is therefore not reported here.

	***um**stellen*		*um**stel**len*	
	first part	second part	first part	second part
duration (ms)	158	360	98	386
pitch (Hz)	170 to 153	110 to 82	101	165 to 82
spectral slope (dB)	-30	-22	-29	-17
intensity (dB)	78	69	70	75

Table 4.1: Acoustic parameters of the first and second part of the two words.

As starting point for the manipulations that we performed in Praat (Boersma & Weenink, 2016), we took the first part plus fricative from the original *umstellen* and the second part from the original *umstellen*, so both parts had stress and clearly articulated segments, and we adjusted this combined sound file to the parameters given below. The vowel quality was not manipulated, but corresponded to the formant values of both stressed vowels in natural speech.

We used the **duration** of the original initial word parts (158 ms for stressed and 98 ms for unstressed) as the two end values on our duration scale, and created a third, middle value at a fractional step of 1.2697, i.e. at 124.4 ms, see the first and second row in Table 4.2. In the second word part (measured from release of the [t]), the two original durations were too close to each other (with less than 20% noticeable difference between them), probably due to phrase-final lengthening. We therefore decided not to vary the duration of the second part but to employ an in-between value of 378 ms for all stimuli.

For the **pitch** manipulation, we used the natural pitch contours of the two words and created an in-between pitch contour with a slight fall from 131 Hz (9.01/2 semitones) to 124 Hz (7.19/2 semitones) for the first word part, and a slight fall from 135 Hz (7.02/2 semitones) to the 82 Hz that both words shared for the second word part, yielding in total three pitch contours, see Table 2, rows three and four.

Duration of first part	158 ms	124.4 ms	98 ms
	= long	= mid	= short
Pitch contour of first and second part	170 to 153 Hz 110 to 82 Hz	131 to 124 Hz 135 to 82 Hz	101 Hz 165 to 82 Hz
	= early peak	= two peaks	= late peak
Spectral slope of second part	-22	-19.5	-17
	= high slope	= mid slope	= low slope
Intensity of first and second part	78 dB 69 dB	74 dB 72 dB	70 dB 75 dB
	= falling	= level	= rising

Table 4.2: Parameters of the manipulation. First rows: actual values, second rows: labels used in the following descriptions. The values in the left column correspond to realizations with stress on the first part, the values in the right column to those with stress on the second part, while the values in the middle column correspond to ambiguous realizations.

With respect to **spectral slope**, the original unstressed and stressed first parts differed only marginally, and therefore only one, intermediate value was taken. For the second word part, we used the values of the original recordings and created an intermediate middle value, resulting in three spectral slope patterns, see Table 2, rows five and six.

In a final step, the **intensity** of the two syllables was manipulated, based again on the measures of the original recordings. In addition to these two, we also created an in-between intensity contour; see rows seven and eight in Table 2. This resulted in a total of 81 stimuli (= 3 duration values * 3 pitch contours * 3 slope patterns * 3 intensity contours).

The two re-synthesized endpoint stimuli were tested in a pilot study with 5 native listeners, to check whether those were consistently categorized, which was the case.

For the EEG experiment, only four of the stimuli from the behavioral task were used. Two of these were the two re-synthesized endpoint stimuli. The choice of the other two was based on the results of the behavioral task, which will be discussed in the results section below. Due to experimental constraints, only the two most natural stimuli were used as standards in the oddball paradigm.

2.2. Participants

Amusics and controls were matched for age, gender, handedness, education and musical training in both studies. All participants were native speakers of German, right-handed and had no self-reported psychological or neurological disorders. They had normal hearing defined as a mean hearing level of 20 dB or less in both ears, assessed by a pure tone audiometry at 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz. The participants were recruited from an existing pool of amusics. Congenital amusia was diagnosed based on the three pitch-based subtests and the rhythm subtest of the MBEA and a detailed questionnaire about their educational and musical background. Only amusics exhibiting both a pitch perception and a rhythm perception deficit were included in this study to ensure homogeneity as much as possible. MBEA scores as the proportion correct out of 30 and *d'* scores on the four subtests that we employed are given below. The difference between the two groups was calculated based on *d'* values. Our amusic group falls below traditional proportion correct cut-off scores from Peretz. et al. (2003) on all but the memory subtest.

All participants were tested in a sound-attenuated chamber in the phonetics laboratory at the University of Düsseldorf. The Ethical Committee of the Medical Faculty at the University of Düsseldorf approved the study protocol, and each participant signed an informed consent form before the experiment, and received a small monetary reimbursement for their participation afterwards.

2.2.1. Behavioral Experiment

10 controls and 7 amusics were included in the behavioral studies. Their characteristics can be found in Table 4.3, and their MBEA scores for the four relevant subtests in Table 4.4. The discriminatory ability of the two groups is significantly different, as shown by the *t*-test values in Table 4.4.

Group		Age	Years of education	Years of music education	Gender
amusic	Mean	27.43	18.86	3.00	2 male
control	Mean	29.00	19.50	2.60	3 male
t-test (df 15)	*t*	-1.179	-0.541	0.398	
	p	0.257	0.297	0.696	

Table 4.3: Subject characteristics: Descriptive statistics and results of *t*-tests comparing amusic (N=7) and control (N=10) participants characteristics. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

			Scale	Contour	Interval	Rhythm	Meter	Memory
Pro-portion Correct	control	Mean	27.50	28.40	28.70	28.50	28.60	28.90
		SD	1.72	1.07	1.25	0.71	1.96	1.20
	amusic	Mean	20.29	20.86	19.71	22.43	22.00	25.57
		SD	3.04	1.21	2.81	2.51	3.06	2.82
d'	control	Mean	3.34	3.67	3.78	3.59	3.84	3.99
		SD	0.83	0.45	0.63	0.39	1.01	0.62
	amusic	Mean	1.23	1.43	1.00	1.60	1.32	2.20
		SD	0.71	0.47	0.43	0.76	0.66	1.07
t-test (df 15)		*t*	-5.50	-9.95	-10.08	-7.10	-5.75	-4.38
		p	0.000	0.000	0.000	0.000	0.000	0.001

Table 4.4: Montreal Battery of Evaluation of Amusia Scores .Means and standard deviations of proportion correct scores (number of correct responses, out of 30), and *d'* scores for amusics (N = 7) and controls (N = 10). The t-test was calculated on *d*'-values. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

2.2.2. MMN Experiment

10 controls and 10 amusics were included in the behavioral studies. Their characteristics can be found in Table 4.5, and their MBEA scores for the four relevant subtests in Table 4.6. Again, the discriminatory ability of the two groups is significantly different, as shown by the *t*-test values in Table 4.6.

Group		Age	Years of education	Years of music education	Gender
amusic	Mean	36.70	19.10	2.40	3 male
control	Mean	33.50	19.00	2.10	3 male
t-test (df 18)	*t*	0.496	0.074	0.252	
	p	0.626	0.942	0.804	

Table 4.5: Subject characteristics: Descriptive statistics and results of *t*-tests comparing amusic and control participants (N= 10 per group) characteristics. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

			Scale	Contour	Interval	Rhythm	Meter	Memory
Pro-portion Correct	control	Mean	27.50	28.30	28.50	28.60	28.50	29.00
		SD	1.72	1.16	1.35	0.84	1.90	1.25
	amusic	Mean	21.60	20.10	19.30	23.50	21.70	24.70
		SD	1.96	2.60	1.34	3.44	4.32	3.06
d’	control	Mean	3.40	3.58	3.66	3.67	3.76	4.07
		SD	0.77	0.56	0.73	0.51	0.97	0.65
	amusic	Mean	1.33	1.10	0.97	1.96	1.45	2.44
		SD	0.48	0.50	0.36	0.93	1.12	1.05
t-test (df 18)		*t*	-7.25	-10.54	-10.45	-5.09	-4.92	-4.18
		p	0.000	0.000	0.000	0.000	0.000	0.001

Table 4.6: Scores for four subtests of the Montreal Battery of Evaluation of Amusia: Means and standard deviations of proportion correct scores (number of correct responses, out of 30), and *d’* scores for amusics and controls (N= 10 per group). The t-test was calculated on *d*’- values. *t*: test statistic of the independent samples *t*-test; *p*: probability value.

2.3. Procedure

2.3.1. Behavioral paradigm

In the behavioral experiment, the 81 stimuli were presented in isolation, as the original recordings on which the stimuli were based on were also words in isolation. The stimuli consisted of three syllables, and therefore provided enough information for the listeners to normalize for speech rate and speaker. Participants were presented with a forced-choice identification task in three blocks with two pictures as answer choices. They heard one stimulus at a time and had to click on one of two pictures in order to answer. Each picture depicted a typical scene for the meaning of the corresponding stress pattern. For *umstellen* "to reposition" two people moving a couch are shown. For *umstellen* "to surround" a group of people surrounding a house are depicted. The two pictures were introduced to the participants before the start of the experiment. The pictures were counterbalanced between blocks and participants. Each stimulus was repeated three times and occurred once per block. Participants took a break after each block. The behavioral task lasted approximately 45 to 60 minutes.

2.3.2. EEG Paradigm

The EEG session was recorded approximately six months after the behavioral session. Each session lasted about 3.5 hours in total, of which about 100 minutes were EEG recording time. Each session consisted of four approximately 25-minute recording blocks with short breaks in between. Participants were watching a silenced nature documentary without subtitles or visible lip movements, while completing a passive listening task. We used nature documentaries to avoid having lip movements of a different language than the one being tested in the video material, as this has been shown to distort EEG results (Kang, Johnson, & Finley, 2016; Shinozaki, Hiroe, Sato, Nagamine, & Sekiyama, 2016). Participants completed a passive listening task as they were instructed to disregard the sounds and to focus their attention on the movie. The auditory stimuli were presented at 60 dB via two loudspeakers placed in front of the participant at a distance of approximately 1 m.

The auditory stimuli were presented in a multi-deviant oddball paradigm with two different blocks. Each block was repeated once, resulting in four blocks in total, which were counterbalanced across participants. A total of 1800 stimuli occurred in each block. The standard stimulus in each block was one of the two stimuli in line with the natural distribution of cues, with all cues indicating stress either on the first syllable (as in *umstellen*) or on the second syllable (as in *umstellen*). The three deviants consisted of the other stimulus with naturally distributed cues and two further stimuli that contained a contradiction between pitch and all other cues, as described in detail in section 2.1 above. This means 4 different stimuli were used. The standard occurred 85 % of the time and each deviant occurred 5 % of the time. Each block started with 20 standards, followed by the oddball sequence, and each deviant was separated from the next by at least 4 and at most 8 standards. The inter-stimulus interval was varied randomly between 300 ms and 500 ms. A variable inter-stimulus interval was chosen to avoid entrainment effects to the stimulus chain (Repp & Su, 2013; Tal et al., 2017). This resulted in 8 event-related potentials (ERPs) per participant: 2 standards and 6 deviants. These 8 ERPs were used to calculate MMNs for 6 different conditions, as the MMN is derived by subtracting the average ERP of the standard from each of the three deviants.

The most negative peak in the 100 to 275 ms post stimulus onset was determined per condition. Each MMN amplitude was calculated as the mean voltage over a 40 ms time window centered at the most negative peak.

2.3.3. EEG Parameters and Pre-processing

The EEG recording and pre-processing parameters are identical to those described in Chapter 3 in section 2.5.

The EEG was recorded using a BioSemi Active Two system (Biosemi Instrumentation BV, Amsterdam, The Netherlands) with 64 Ag-AgCl electrodes that were placed according to the international 10/20-system in a cap fitting the participant's head size. 7 further electrodes were placed on the tip of the nose, the left and right mastoid, below and above the left eye and the outer canthi of the left and right eyes (recording the electro-oculogram; EOG). The EEG signal was recorded at 8192 Hz and later down-sampled to 512 Hz. The subsequent analyses were performed in Praat (Boersma & Weenink, 2016). The data were offline referenced to the average of the two mastoid channels. Slow drifts were removed by subtracting a line from each channel so that the first and the last sample become zero. The data were bandpass filtered in the frequency domain with a low cut-off of 1 Hz (0.5 Hz bandwidth) and a high cut-off of 25 Hz (12.5 Hz bandwidth). The EEG was segmented into epochs of 500 ms, from 110 ms before to 390 ms after stimulus onset. For baseline correction, the mean voltage of the 110 ms pre-stimulus served as a baseline for amplitude measurement and was subtracted from each sample in this epoch. Artifact correction was done automatically and epochs with an EEG or EOG change exceeding +/-75 mV were excluded. Participants with more than 50% of artifact-contaminated epochs would have been excluded from analysis. No participant exceeded this limit. This way at least 90 deviants per type remained in the analysis.

3. Results

3.1. Behavioral

A generalized linear mixed effects model with subjects and stimulus item as random effects and everything else as fixed effects (i.e. the complete model) was calculated. The number of answer to “stress on the first word part” was taken as the dependent variable. Following established modeling practices (Baayen, Davidson, & Bates, 2008), a full model that included all variables and interactions was used as a starting point. Stimuli were contrast coded and only explicit contrasts were used, i.e. variables were centered around zero. The regression models were then simplified by the stepwise exclusion of non-significant fixed effects. A fixed effect was considered non-significant if its *p*-value was higher than 0.05, if there were no significant interactions including it and if the Akaike Information Criterion of the model including the predictor was higher than when the predictor was not included. The final model only contained group, pitch and duration as fixed effects, as intensity and slope did not yield any significant results. We found main effects of duration and of pitch and an interaction between duration and group. The model specifics can be found in Table 4.7. To understand the Group by Duration interaction, separate models with only duration as fixed factor were performed for each group separately. This revealed an effect of duration in the amusic group (B = -1.75, SE = 0.79, $z = 2.20$, $p = 0.028$), but not in the control group (B = 0.74, SE = 0.76, $z = 0.98$, $p = 0.33$).

	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-0.76	0.16	-4.73	2.17e-06 ***
Group	-0.25	0.18	-1.35	0.17801
Duration	1.15	0.35	3.24	0.00119 **
Pitch	5.37	0.36	14.99	< 2e-16 ***
Group by Duration	-0.85	0.28	-3.01	0.00217 **
Group by Pitch	-0.17	0.30	-0.57	0.56916
Duration by Pitch	0.42	0.88	0.48	0.63341
Group by Duration by Pitch	-0.31	0.74	-0.42	0.67724

Table 4.7: Final generalized linear mixed model fit by maximum likelihood. Asterisks indicate significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

A visualization of the data can be found in Figures 4.1 and 4.2. Figure 4.1 depicts a clear effect of pitch: Both amusics and controls identified an early pitch rise as stress on the first syllable and a late pitch rise as stress on the second syllable, while the ambiguous stimuli with two pitch peaks were mainly identified as having stress on the second syllable.

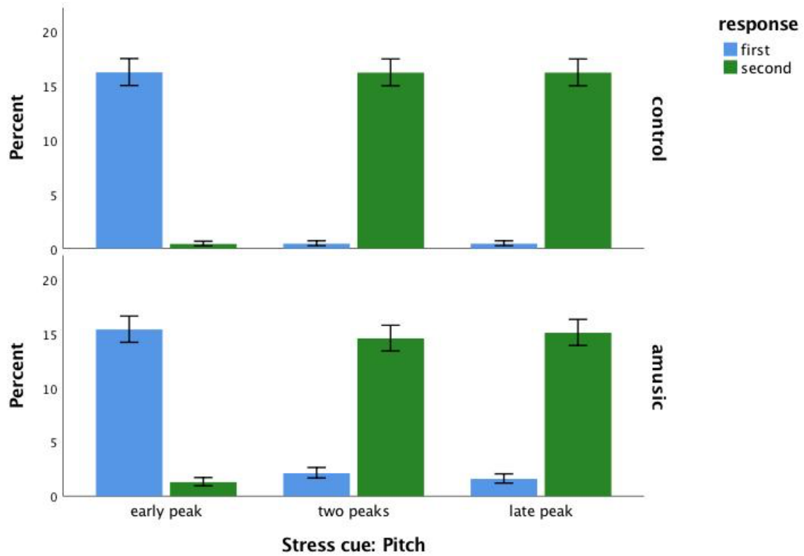


Figure 4.1: Response first (blue) or second syllable (green) in percent (grand total) to stimuli where pitch has an early peak, two peaks, or a late peak. Top: control group, bottom: amusic group. Error bars show 95% CI.

Figure 4.2 depicts the effect of duration. All participants showed a preference to identify stimuli as being stressed on the second syllable. However, while the responses of the control group (upper panel) are not influenced by the duration of the first vowel (the ratio of their answers stays the same across all three vowel durations), the responses by the amusic group (lower panel) depend on the duration of the vowel, with far more "stress on first syllable" identifications for stimuli with a long than with a mid or short first vowel. The amusics thus seem to use durational cues more reliably than controls to identify stress, indicating a better performance for amusics.

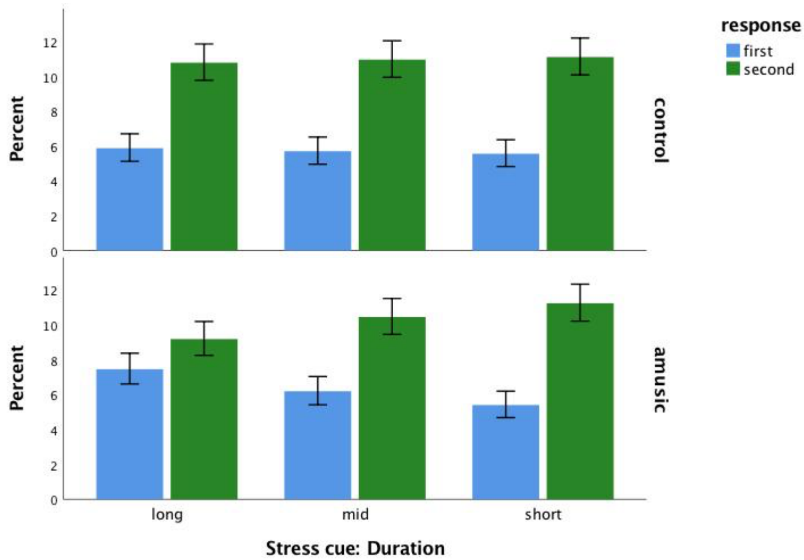


Figure 4.2: Response first (blue) or second syllable (green) in percent (grand total) to long, mid or short duration of the first vowel. Top: control group, bottom: amusic group. Error bars show 95% CI.

Due to non-significant results for slope and intensity, only two stimuli that contained a contradiction between pitch and all other cues were further selected for the EEG study: the first had an ambiguous pitch contour (two peaks) whereas all other parameters corresponded to those for stress on the first word part, and the second had an early peak pitch contour (corresponding to stress on the first part) whereas all other parameter settings corresponded to stress on the second part.

3.2. EEG

The EEG analysis was run on the MMN amplitude measured at 9 channels (Fz, FCz, Cz, F3, F4, FC3, FC4, C3, C4).

Visual analysis of the scalp topography confirms the negative polarity, the expected latency and fronto-central scalp distribution of the MMN for the controls, cf. Figure 4.3. Amusics overall do not show a strong negativity, as indicated by the lighter blue color compared to the control group in the left panel. The right panel shows the average difference waveform averaged across all conditions for amusics (dotted line) and controls (solid line).

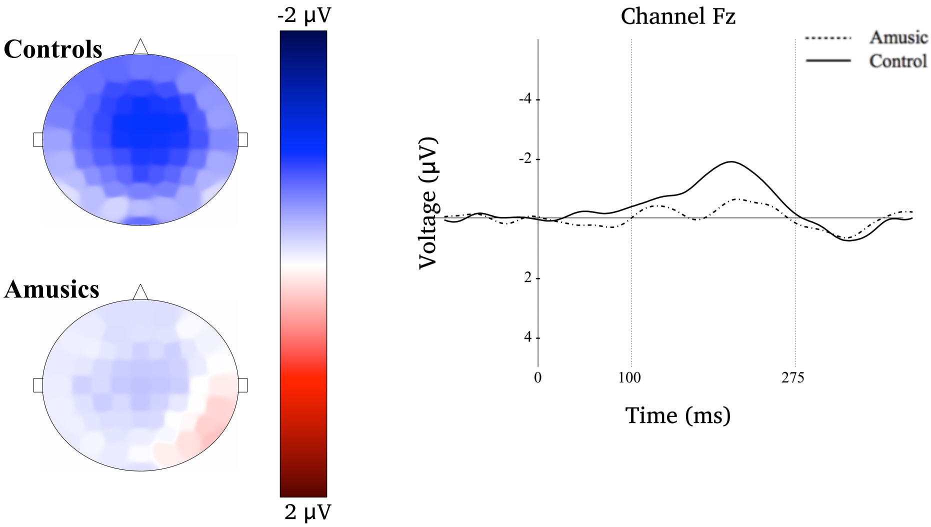


Figure 4.3: Difference curves (deviant - standard) for amusics and controls in a time window between 100 and 275 ms averaged across all conditions. Left panel are the topographical maps and right panel the grand average difference waves plotted at Fz.

Figure 4.4 shows the topographical plot per condition and group. The dark blue color for the topographical plot on the upper left indicates that the natural deviant stressed on the second syllable with the standard stressed on the first elicited the strongest MMN in controls. A similar pattern, though far less strong, can be found for amusics, see second plot in the left column. The reverse condition with deviant stressed on the first and standard on the second syllable elicited quite a strong MMN again in controls, but not so in amusics, cf. plot 3 and 4.

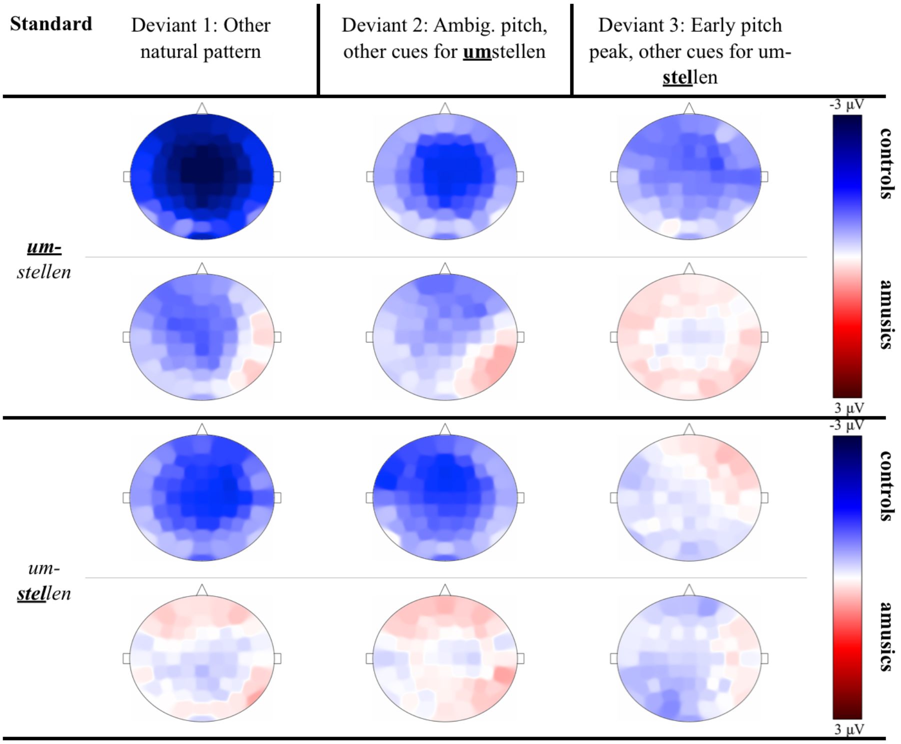


Figure 4.4: Topographical maps of amusics and controls averaged in a time window between 100 and 275 ms per condition. Controls at the top and amusics at the bottom per condition.

After visual analysis, we performed two-tailed *t*-tests against zero separately for each group to determine whether the difference waveform response was present in every condition, all of which were significant, see Table 4.8.
A laterality analysis was carried out but revealed no significant differences between conditions or groups.

Standard	**Deviant 1:** other natural pattern	**Deviant 2:** ambiguous pitch, other cues for *umstellen*	**Deviant 3:** early pitch peak, other cues for *umstellen*
umstellen	-4.51*** -2.79***	-2.67*** -1.73***	-2.81*** -1.96***
umstellen	-2.34*** -1.97***	-2.09*** -1.84 ***	-1.05*** -1.28***

Table 4.8: Mean voltage at electrode Fz measured in μV. Top value is that of the control group, bottom value that of the amusic group. * indicate significance levels in *t*-tests against zero: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

As we were interested in the differences between amusics and controls in every condition, we calculated a linear mixed model (lmer) in R with subject as random effect and group and condition as fixed factors. Coding and modeling were carried out similar to the description in section 3.1 above. A summary of the model can be found in Table 4.9.

	Estimate	**Std. Error**	***t*-value**	***p*-value**
(Intercept)	-3.23	0.19	-17.37	<0.001
Group	-1.68	0.37	-4.52	<0.001
Condition	0.32	0.02	15.56	<0.001
Group by Condition	0.34	0.04	8.21	<0.001

Table 4.9: Summary of the model.

We found a main effect for group with amusics (M = -1.93) overall having a smaller MMN than controls (M = -2.58), a main effect for condition, and an interaction between group and condition. To understand the Group by Condition interaction, separate t-tests were performed for each condition separately. They all yielded significant results, with controls having a more negative MMN than amusics, except in one case: When the standard was stressed on the second syllable and Deviant 3 from Table 4.8 was used, then there was no significant effect, also depicted in the bottom right topographical maps in Figure 4.4.

4. Discussion

Our experiments tested the identification of word stress by amusics and controls based on the acoustic cues of pitch, duration, intensity and spectral slope. Spectral slope and intensity did not have a significant effect on identification.

To our surprise, both controls and amusics used pitch in a very similar way for the identification of word stress: Unambiguous pitch cues were used to identify the corresponding word stress patterns, while the ambiguous stimuli with two pitch peaks were mainly identified as having stress on the second syllable. Though there were small differences in the behavior of the two groups, these were not significant. This finding is unexpected as amusics were expected to show difficulties in their pitch perception.

Even more surprising was the finding that amusics seemed to use durational cues more reliably than controls to identify stress on the first syllable. This finding could hint at compensation strategies that amusics may have developed to compensate for their pitch perception deficits.

In addition, amusics did display a significantly reduced MMN in comparison to controls in all but one condition: When the stress of the standard was on the second syllable and the deviant had a pitch peak on the first syllable while all other cues indicated stress on the second syllable, there was no significant difference between amusics and controls. However, the MMNs were still present in both groups.

These findings taken together are surprising, as amusics seem to be able to identify something in the behavioral task to which they show an at least reduced early neural response. It seems to stand to reason that amusics might compensate for their reduced early change detection responses at later processing stages. Our findings are somewhat comparable to those by Jasmin et al. (2019) who found no differences between amusics and controls in their phonetic cue weighting but did find that amusics placed greater emphasis on durational cues than on pitch cues in their prosodic cue weighting task. Further investigations are needed to untangle whether Jasmin et al.'s (2019) findings concerning durational differences in consonants (VOT), i.e. that there was no difference in the cue usage between amusics and controls, also hold for vowels, or whether vowel duration would be more comparable to their and our findings concerning word stress, i.e. that amusics placed greater emphasis on durational cues than on pitch cues. Regardless of this, their findings also point to a possible compensation strategy of amusics.

Our ERP results show that the auditory system of amusics has formed a representation of the repetitive aspect of the standard stimulus as represented by the MMN that they exhibit. However, their MMN is significantly reduced in comparison to our control population and might therefore represent a more inaccurate discrimination of the stimuli than the controls' larger MMN does (Kujala & Näätänen, 2001). This finding is in line with Weber et al.'s (2005) finding of infants at risk of SLI showing a significantly reduced MMN to changes in stress patterns. A further comparison of amusia to other developmental disorders and their behavioral and neurophysiological markers seems warranted.

The general finding that amusics did indeed display an MMN response is in direct opposition to that by Braun et al. (2008), who did not find an MMN in amusics at all. Furthermore, the fact that the response of our amusics was significantly reduced is in opposition to the findings by Moreau et al. (2009; 2013) and Mignault et al. (2012) who found that amusics displayed completely normal MMNs. All these findings utilized musical stimuli, however. Our study supports the results by Nan et al. (2016), who found a reduced MMN for tone-language speaking amusics in response to lexical tones.

Our study, as any study, can be criticized in a number of ways: Firstly, concerning our methodology: The small sample size can be seen as problematic for an ERP study. This is a pitfall that all studies with amusics face, as the population is a rather small one. However, our sample size is fairly large in comparison to other ERP studies with amusics. Secondly, the choice of our stimuli could also be criticized. Due to time constraints, we had to limit our study to one stress minimal pair and its manipulations, even though we piloted many more. We would have liked to include more items; this, however, proved to be impossible without making the actual experiment last much too long. So we decided to focus on one specific pair and to focus on our manipulations. Lastly, precisely these can also be criticized. The phonetic manipulations and acoustic correlated for stress, even though carefully based on previous research, are always debatable.

5. Conclusion

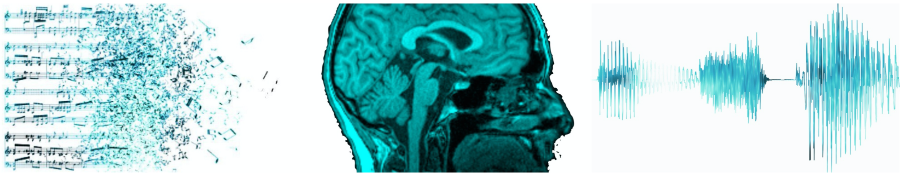
Our study is the first to investigate word stress processing in congenital amusia. By employing a behavioral and electrophysiological paradigm, we were able to assess both in relation to each other. Our results surprisingly show that amusics utilize durational cues more than controls and that they did not struggle with pitch cues when it comes to the identification of word stress.

We were also able to demonstrate that amusics did indeed show an MMN, as an automatic reaction to auditory change detection to different word stress patterns. However, this MMN was significantly reduced in comparison to our control population, indicating abnormal neural processes even at this very early stage of processing.

Taken together, these findings show that amusics exhibit difficulties when it comes to word stress processing, but might have developed compensation strategies. Further studies investigating later ERP components, as well as different linguistic manipulations, and a comparison to other developmental disorders are therefore warranted.

6. Acknowledgements

We thank Dirk Vet for helping us writing the E-Prime script used in the EEG study.



5. The Nature and Nurture of Congenital Amusia: A Twin Case Study

<u>This chapter is based on:</u>

Pfeifer, Jasmin & Silke Hamann (2018). The Nature and Nurture of Congenital Amusia: A Twin Case Study, *Frontiers in Behavioural Neurosciences 12: 120*.

Abstract

In this paper we report the first documented case of congenital amusia in dizygotic twins. The female twin pair was 27 years old at the time of testing, with normal hearing and above average intelligence. Both had formal music lesson from the age of 8 to 12 and were exposed to music in their childhood. Using the Montreal Battery of Evaluation of Amusia (Peretz et al., 2003), one twin was diagnosed as amusic, with a pitch perception as well as a rhythm perception deficit, while the other twin had normal pitch and rhythm perception.

We conducted a large battery of tests assessing the performance of the twins in music, pitch perception and memory, language perception and spatial processing. Both showed an identical albeit low pitch memory span of 3.5 tones and an impaired performance on a beat alignment task, yet the non-amusic twin outperformed the amusic twin in three other musical and all language related tasks. The twins also differed significantly in their performance on one of two spatial tasks (visualization), with the non-amusic twin outperforming the amusic twin (83% vs. 20% correct). The performance of the twins is also compared to normative samples of normal and amusic participants from other studies.

This twin case study highlights that congenital amusia is not due to insufficient exposure to music in childhood: The exposure to music of the twin pair was as comparable as it can be for two individuals. This study also indicates that there is an association between amusia and a spatial processing deficit (see Douglas & Bilkey, 2007; contra Tillmann et al., 2010; Williamson et al., 2011) and that more research is needed in this area.

1. Introduction

Congenital amusia is an innate disorder that has been shown to have a negative influence on pitch perception (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Peretz et al., 2002; Stewart, 2008), with a co-occurring deficit in rhythm perception in about 50% of the cases (Pfeifer & Hamann, 2015). This congenital variety of amusia is neither caused by a hearing deficiency nor by any form of brain damage or intellectual impairment (Ayotte, Peretz, & Hyde, 2002) and causes persistent, lifelong impairments in the musical (Stewart, 2008), or more broadly, auditory domain. While congenital amusia had long been reported to affect only the musical domain (Ayotte et al., 2002; Peretz, 2001; Peretz et al., 2002), many recent studies have shown that different areas of speech perception are also affected, such as the perception of intonation (Hamann, Exter, Pfeifer, & Krause-Burmester, 2012; Liu, Patel, Fourcin, & Stewart, 2010; Patel, Wong, Foxton, Lochy, & Peretz, 2008), of tone in languages that employ tone differences distinctively (Liu, Jiang, Wang, Xu, & Patel, 2015; Liu, Maggu, Lau, & Wong, 2015; Liu, Xu, Patel, Francart, & Jiang, 2012; Tillmann et al., 2011), the perception of vowels (Huang, Zhang, Shi, Yan, & Wang, 2016; Zhang, Shao, & Huang, 2017) and of emotional prosody in language (Lolli, Lewenstein, Basurto, Winnik, & Loui, 2015; Thompson, Marin, & Stewart, 2012).

The prevalence of the disorder is estimated to range between 1.5% (Peretz & Vuvan, 2017) and 4% (Kalmus & Fry, 1980) of the general population. Because of its clustering within families, documented in the first and so far only familial aggregation study by Peretz, Cummings, & Dube (2007), congenital amusia has been hypothesized to have a genetic component. Peretz et al. studied 23 amusics from 9 families and calculated a sibling recurrence risk ratio (the ratio of manifestation, given that a sibling is affected, compared with the prevalence in the general population; Risch, 1990) of λ_s =10.8. This ratio is in the same order of magnitude as the heritability of specific language impairments and of absolute pitch. Based on these numbers, recent studies think it very likely that congenital amusia has a hereditary component (Gingras, Honing, Peretz, Trainor, & Fisher, 2015; Peretz et al., 2007; Peretz & Vuvan, 2017). However, familial aggregation could be simply due to shared family environment (in the case of congenital amusia, e.g. non-exposure to music within a family). Such environmental factors can only be reliably separated from genetic effects in twin studies, which have been employed successfully to test the heritability of pitch processing in general. Drayna, Manichaikul, de Lange, Snieder, & Spector (2001), for instance, compared musically non-preselected monozygotic (N=136) and dizygotic (N=148) twin pairs using the Distorted Tunes Test (DTT; Kalmus & Fry, 1980) in a large-scale study. The heritability of pitch processing as estimated by their genetic model fitting was 71%, and they found a high correlation (0.67) in liability within monozygotic twin pairs and a medium one (0.44) for dizygotic twin pairs. A newer twin study on general pitch and rhythm perception (Seesjärvi et al., 2016) used three subtests from an online musicality test (Peretz et al., 2008) with 69 monozygotic and 44 dizygotic twin pairs to compare genetic and environmental effects. The correlations of scores within the twin pairs on the scale test was comparable to Drayna et al.'s (2001) with a high correlation (0.58) for monozygotic and a medium one (0.38) for dizygotic twin pairs. On the out-of-key test, a high correlation was found for both twin groups (0.63 monozygotic and 0.67 dizygotic) and on the off-beat test only a medium correlation (0.31) for monozygotic twin pairs. Mosing, Pedersen, Madison & Ullen (2014) tested a large sample of 2568 Swedish twins with a rhythm, a melody and a pitch task. They also found similarly high correlations of 0.57 for melody and 0.48 for pitch in monozygotic twins but lower correlations of 0.32 for melody and 0.29 for pitch in monozygotic twins.

A pitfall of utilizing such twin studies in congenital amusia research is the sample size. The recruitment of amusic participants in general is already difficult, while the recruitment of a sufficiently sized pool of amusic twin pairs is nearly impossible. Most amusia studies have small sample sizes, and some are single subject studies, e.g. Peretz et al. (2002) reporting the first case of amusia or Lebrun, Moreau, McNally-Gagnon, Mignault Goulet, & Peretz (2012) reporting the first case of amusia in a child.

In the present study, we report the first documented case of congenital amusia in a dizygotic twin. With these twins, we conducted a large battery of tests assessing their musicality, pitch perception and pitch memory, language perception, and spatial abilities in order to determine a possible genetic impact of amusia on these abilities. An overview can be found in Table 5.1. We chose to use not only the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003) to assess amusics' music perception, as it has been criticized lately (Henry & McAuley, 2013; Pfeifer & Hamann, 2015) but also conducted the Goldsmith Musical Sophistication Index (GoldMSI, Müllensiefen, Gingras, Musil, & Stewart, 2014). The Gold-MSI has never been conducted with amusics to our knowledge, so our twin pair will be compared to available norm samples. We thereby hope to obtain a broader perspective on the musical abilities and disabilities of our amusic twin in comparison to the non-amusic twin. We also included pitch perception tasks, as these are now widely used to determine amusics' pitch thresholds, and memory span tasks to investigate possibly different memory capacities of the twins. In addition, we wanted to asses the twins' language perception, as an increasing body of literature points to deficits in speech perception as well. We decided to also include tests on spatial abilities, as deficits in spatial processing by amusics have been found by Douglas & Bilkey (2007). Douglas and Bilkey used a classic Mental Rotation task (Shepard & Metzler, 1971) with line drawings of two three-dimensional objects that had to be compared, and amusics showed significantly higher error rates on this task. Later tests failed to replicate these findings. Tillmann et al. (2010) utilized the same Mental Rotation task but with 160 trials instead of the 20 employed by Douglas & Bilkey. In addition, they also used a bisection task in which the midpoint of a straight line or a string of letters has to be marked. They found no difference between controls' and amusics' accuracy or reaction time on either task. Williamson, Cocchini, & Stewart (2011) again used a version of the Mental Rotation task and two further tasks assessing memory for sequences of spatial location (Milner, 1971) and memory for visual patterns (Della Sala, Gray, Baddeley, & Wilson, 1997). No difference in accuracy between amusics (N=14) and controls (N=14) on any of these tasks was found. However, a subgroup of amusics with the most severe pitch perception deficits exhibited slower reaction times on the Mental Rotation task. Peretz and colleagues (Peretz et al., 2008; Peretz & Vuvan, 2017) report that amusia and visuo-spatial deficits are associated, though this is solely based on self-report questionnaire data.

2. Materials and Methods

2.1. Procedure

First we assessed the twins with the Montreal Battery of Evaluation of Amusia (Peretz et al., 2003) and a questionnaire about educational, musical and demographic background. In addition, we assessed the twins' hearing and their intelligence. In order to further ascertain the differences and similarities in their musical, pitch perception and memory, language and spatial abilities, we then conducted a number of additional tests, listed in Table 5.1.

All experiments were conducted at the University of Düsseldorf in the phonetics laboratory in a sound-insulated booth. All experiments were programmed in Praat (Boersma & Weenink, 2016) unless otherwise mentioned, and auditory stimuli were presented over AKG K 601 headphones on a windows XP computer. All data were collected in accordance with the declaration of Helsinki. Both participants gave informed written consent to participate in this study and received a small monetary reimbursement for their time. Both participants completed all test over the course of several days. The twins took the same tests on the same days, right after each other so that they did not have the possibility to exchange information on the tasks before both had completed them.

Ability	Task	Subtests	Reference
Musical	Goldsmith Musical Sophistication Index (Gold-MSI)	Questionnaire	Müllensiefen, Gingras, Musil, & Stewart, 2014; Fiedler & Müllensiefen, 2015; Schaal, Bauer, & Müllensiefen, 2014
		Gold-Genre	
		Gold-Melody	
		Gold-BAT	
Pitch perception and memory	Pitch perception task	Detection	Williamson & Stewart, 2010
		Direction	
	Memory span task	Pitch Span	Williamson & Stewart, 2010; Schaal, Pfeifer, Krause, & Pollok, 2015
		Visual Span	
Language perception	Intonation task	Intonation	Hamann, Exter, Pfeifer, Krause-Burmester, 2012
	Vowel perception task	Vowel	Pfeifer & Hamann in prep.
Spatial	Object Perspective Taking Test	Orientation	Hegarty & Waller, 2004
	Santa Barbara Solids Test	Visualization	Cohen & Hegarty, 2012

Table 5.1: Overview of the assessed abilities and the utilized tests with references.

2.2. Participants

The female twins were 27 years old at the time of testing with no history of psychiatric or hearing disorders. They grew up together in the same household with one younger male sibling and attended primary and secondary school and their undergraduate program in linguistics together. They had music lesson (flute) from the age of 8 to 12 and had the same exposure to music in their childhood and adolescence. The parents of the twins still live together.

The mother does not show signs of amusia and seems to enjoy music. The father, however, has a severe hearing deficit in both ears that has been present since childhood due to a measles infection, and he uses hearing aids. He therefore had no normal exposure to music in childhood. Due to his severe hearing impairment, we could not test him for amusia and we cannot make any statement whether he might be amusic or not.

For the diagnosis of the twins, the MBEA (Peretz et al., 2003) and a questionnaire were used (the latter is described in detail in Pfeifer & Hamann, 2015: 9–11). Their scores on the MBEA are given in Table 5.2.

	Scale	**Contour**	**Interval**	**Rhythm**	**Meter**	**Memory**
sum correct responses A	20 (22)	21 (22)	20 (21)	21 (23)	25 (20)	26 (22)
sum correct responses C	27 (22)	26 (22)	26 (21)	29 (23)	24 (20)	28 (22)
d′ A	1.8	2.07	1.25	1.42	1.95	2.34
d′ C	3.07	2.95	2.95	3.83	1.68	3.44

Table 5.2: MBEA scores of the twins based on sum of correct responses out of 30 where cut-off score by Peretz et al (2003) are given in brackets, and *d′* scores.

One twin, called A in the following, falls below the cut-off scores by Peretz et al. on the first four subtests, exhibiting a pitch and a rhythm perception deficit. The other twin, called C in the following, stays well above the cut-off scores on all subtests. A further analysis of the MBEA results with Signal Detection Theory (SDT) (Green & Swets, 1966; Macmillan & Creelmann, 2005) was carried out, as the SDT measure *d′* is bias free and reflects participants' discriminatory ability without the response bias. The twins show clearly distinct discriminatory abilities, with C having much higher scores i.e. being able to discriminate much better between stimuli than A in all but the Meter subtest, where A is slightly better than her non-amusic twin sister. The *d′* scores for the Meter subtest are rather low for both twins, which reflects the problematic nature of this subtest (see Pfeifer & Hamann, 2015, for details).

The answers to the questionnaire confirmed the results obtained by the MBEA.

Both twins have normal hearing defined as a mean hearing level of 20 dB or less in both ears (tested with a pure tone audiometry at 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz).

The twins intelligence was assessed using the German version of the Hamburger Wechsler Adult Intelligence Scale (HAWIE; Wechsler, 1964). The twins both exhibited higher than average intelligence scores belonging to the highest 2% of scores. The non-amusic twin C achieved a global score of 132 IQ points (verbal 111, action 139) and the amusic twin A a global score of 138 (verbal 124, action 136) IQ points. Both reached similar scores on all subtests with the exception of the digit span subtest, where A had problems in comparison to her twin.

2.3. Further Musical Abilities

In addition to the MBEA, we also employed the Goldsmith Musical Sophistication Index (Müllensiefen et al., 2014), to further assess the musical performance of our twin pair. We tested them with four of the five parts of the Gold-MSI: A self-report questionnaire (the German version hereof, see Fiedler & Müllensiefen, 2015; Schaal, Bauer, & Müllensiefen, 2014), a genre sorting task (Gold-Genre), a melody memory task (Gold-Melody), and a beat alignment perception task (Gold-BAT). The Gold-Genre task consists of 16 musical excerpts, each 800 ms long, without lyrics or vocals. The excerpts are taken from four different genres (pop, rock, jazz and hip-hop) and participants have to group them into four categories without being told what the categories are. The Gold-Melody task consists of 13 melody pairs that have to be compared. Each melody is between 10 and 16 notes long, and the second melody is always transposed to a different key to test memory for a melody's interval structure rather than absolute pitch. The two melodies are either the same - except for the key transposition of which subjects are informed - or the second melody contains an alteration. The Gold-BAT task is based on the Beat Alignment Test by Iversen & Patel (2008) and investigates beat-based processing. The test consists of 12 melodies from three different genres, and a beat track is superimposed on every melody. The participant’s task is to judge whether the beat track is on the beat of the music or not.

2.4. Pitch Perception and Pitch Memory Abilities

We employed two tasks previously used by Williamson & Stewart (2010) to investigate the auditory pitch perception abilities of our two participants. The pitch detection task measures the threshold for the detection of a pitch change, while the pitch direction task measures the threshold for discriminating pitch direction. Both are two-alternative forced choice AXB tasks employing an adaptive two-up-one-down staircase procedure. Every trial consisted of three consecutive tones, each 600 ms long. In the pitch detection task, the target tone was a pitch glide centered around 500 Hz, while the two non-target tones were steady-state tones with a frequency of 500 Hz. In the pitch direction task, all three tones were pitch glides centered around 500 Hz. The target tone was a glide in the opposite direction to the two non-target tones. The task was to identify which tone was different: The first or the last. Each task started with a pitch difference of six semitones. When participants gave two consecutive correct answers, they advanced a level, and the pitch difference became smaller. When they made one mistake, they went one level down and the pitch distance became larger. Each task ended after 15 level changes. To increase the precision of threshold determination, variable pitch step sizes were used. For the first five level changes, the change consisted of one semitone. For level changes 6–9, a change of 0.2 semitones was used, and for levels 10–15 a change of 0.05 semitones. The last 10 trials were averaged to compute the perceptual threshold of the participants.

We also included a test assessing participants' short-term memory for auditory as well as visual sequences with a two-alternative forced choice design (Schaal, Pfeifer, Krause, & Pollok, 2015; Williamson & Stewart, 2010). The auditory stimuli were ten sine wave tones with a duration of 500 ms and with fundamental frequencies ranging from 262 to 741 Hz in whole tone steps. The visual stimuli were ten Devanagari letters presented for 500 ms in black on a white background. The procedure was the same for both types of stimuli: 500 ms of silence or a blank screen were followed by two successive, equally long sequences of tones or letters. The two sequences in a trial were either identical or the position of two tones/visual signs was switched in one of the sequences. The participants' task was to determine whether the two sequences were identical or different. The same two-up-one-down staircase procedure described above for the pitch perception thresholds was employed, and the difficulty advanced, i.e. the sequences became longer after two consecutive correct answers and shorter after one incorrect answer. Each task was terminated after four incorrect answers. The last ten trials of each task were used to calculate participants' memory span, indicating the (auditory or visual) memory load they can store in each domain.

2.5. Language Perception Abilities

2.5.1. Intonation Perception

To test the intonation perception of our twin pair, we used the AX same-different discrimination task and stimuli from Hamann et al. (2012), which was in turn based on the study by Patel et al. (2008). The stimuli pairs were based on recordings of four German statement-question pairs spoken by a male native speaker. Each pair was identical but for the final intonation contour, i.e. statements exhibiting a falling pattern and the corresponding echo questions a rising pattern. The intonation contour of questions was manipulated downwards in seven steps of one semitone each, while the intonation contour of questions was manipulated upwards in the same way. Stimulus pairs consisted of the original statement or question followed either by one of the downward or upward manipulations or the original again, resulting in 112 stimuli pairs. Participants had to indicate for each pair whether the two were identical or not. We also included sinusoidal wave analogs (similar to Patel et al.) that did not contain any linguistic material but were solely based on the intonation contour of the speech stimuli. These were manipulated and paired in the same way as the speech stimuli. The test was scored by calculating three different performance measures: Hit rate, percentage correct and d'. Hit rate is solely based on answers to stimulus pairs where A differs from X, which are considered a hit when they are correctly identified as different. Percentage correct is the sum of both hits and correct rejections (stimulus pairs where A and X are the same and which are correctly identified as same) in relation to all answers.

2.5.2. Vowel Perception

The second language-related task consists of an AXB forced-choice discrimination task with vowel stimuli, identical to the one used in Chapter 2. We used isolated synthetic vowels based on auditory properties of the natural German vowels /ɛ/ and /e:/, where /e:/ is 110 ms long with a first formant (F1) of 350 Hz and a second formant (F2) of 2157 Hz, and /ɛ/ 60 ms long with an F1 of 524 Hz and an F2 of 1869 Hz (based on Jessen 1993).

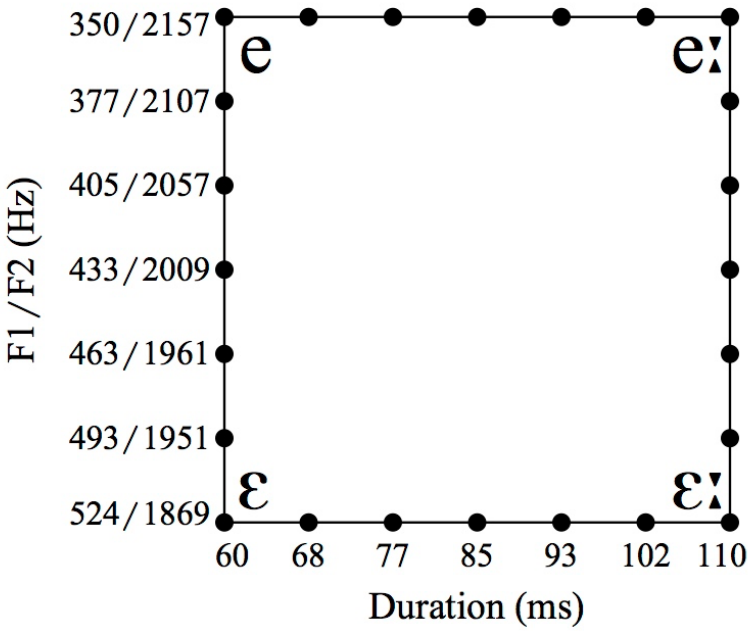


Figure 5.1: Spectral and durational values of vowel stimuli.

On the basis of these vowels we created four continua with seven steps each, depicted as the four sides of the rectangular in Figure 5.1. For each AXB trial, A and B were the endpoints of one continuum (one side of the rectangular), and X could either be one of the two endpoints or one of the five vowels in-between. The trials were offered with two different inter-stimulus intervals of either 0.2 s or 1.2 s (Werker & Logan, 1985; Williamson & Stewart, 2010). Each trial was repeated 5 times throughout the experiment.

The vowel perception task was scored by calculating the percentage of how often participants perceived X correctly as category A (where the answer was considered correct when X was either identical to A or one of the three stimuli close to A on the continuum in question). Based on this measure we calculated *d′* values.

2.6. Spatial Abilities

The Mental Rotation task used in previous studies to test amusics' spatial abilities has been argued to be rather complex and to rely on different cognitive processes (Williamson et al. 2011), we therefore decided to employ the Object Perspective Taking Test (Hegarty & Waller, 2004) and the Santa Barbara Solids Test (Cohen & Hegarty, 2012) instead. These two tests were chosen as they differentiate between spatial orientation abilities, tested with the Object Perspective Taking Test, and spatial visualization abilities, tested with the Santa Barbara Solids Test.

In the Object Perspective Taking Test, the participant is asked to imagine the degree in which several objects are placed to each other from different perspectives, providing a test of egocentric spatial transformations. The test was administered in a paper-and-pencil based version and contained 12 items. Each item consists of a map in the top half of the page, in which seven items are arranged. Participants are asked to imagine being at the position of one object, facing a second one, and having to point to a third object. On the bottom of the page is a circle and the first object is always located in its center with an arrow pointing vertically up to the second object. Participants have to draw a second arrow from the center of the circle outwards to the position of the third target object, thereby making an egocentric transformation. Participants are prevented from rotating the paper, so as not to make the task easier. The perspective change on every item is at least 90 degrees. Each item is scored by calculating the deviation from the correct direction in degrees. The overall score on the test is the average deviation across all items.

The Santa Barbara Solids Test was also administered in a paper-and-pencil version containing 30 items. Each item consists of a three-dimensional geometric object that is sliced by a plane. Participants are asked to imagine looking at the two-dimensional cross-section of the geometric object caused by the plane. The stimuli vary in complexity along two factors: Complexity of the geometric shape and the orientation of the cutting plane. Half of the items have planes that are vertical or horizontal to the main axis of the shape, and the other half have planes that are diagonal to this axis. Participants are given four answer choices, depicted as possible cross-sections. The answers include one egocentric distracter that represents the shape that a participant who fails to change her perspective would choose, providing a way to differentiate whether a perspective change away from egocentric was made or not; see the example in Figure 5.2.

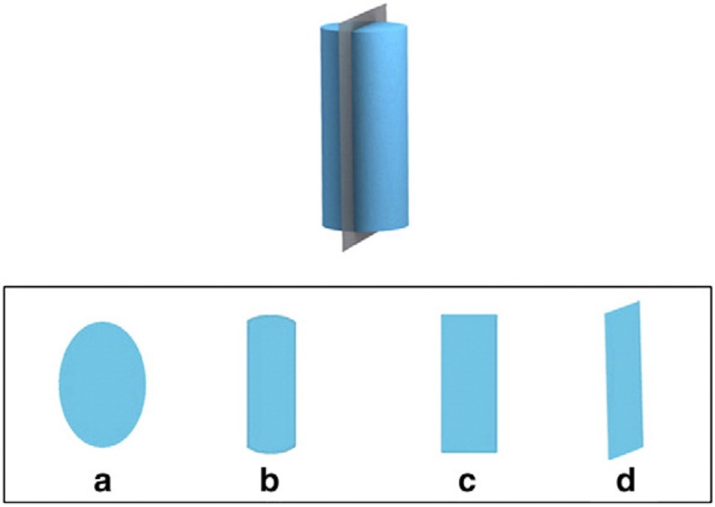


Figure 5.2: Example item from the Santa Barbara Solids Test (Cohen & Hegarty 2012: 869).

The top depicts a three-dimensional object and a plane cutting this object vertically, the bottom displays four cross-sections as answer choices ((c) being the correct answer, and (d) the distracter without change in view perspective). The Santa Barbara Solids Test is scored by counting the number of correct responses and calculating the percentage correct.

3. Results

The pitch perception and memory tasks as well as the language perception tasks have previously been used with amusics, and the performance of the twin pair is compared to those samples. The Gold-MSI has never been conducted with amusics, therefore no cut-off scores for amusics are available. However, Müllensiefen et al. (2014) provide data norms based on 147,636 participants, to which we compared our two subjects. Similarly, the spatial tasks have not been administered to amusics before, and we compared the twins' performance to the data norms by Hegarty & Waller (2004) for the

Object Perspective Taking Test (based on 62 participants) and the norms by Cohen & Hegarty (2012) for the Santa Barbara Solids Test (223 participants).

3.1.1. Further Musical Abilities

The Gold-MSI questionnaire yields six factors, which can be found in Table 3 with mean scores and endpoint of scales from Müllensiefen et al. (2014). A's scores are almost exclusively situated in the lowest percentile in all factors but the musical education one, due to both twins having had music lessons for four years. C's scores are higher and range between the 2nd and 40th percentile.

	Active Engagement	Percetual Abilities	Musical Training	Singing Abilities	Emotions	General Musical Sophistication
A	9 (1)	22 (1)	11 (11-13)	8 (1)	12 (1)	21 (1)
C	19 (2)	48 (37-40)	12 (14-15)	24 (18-19)	32 (26-31)	51 (8)
Scale Min	9	9	7	7	6	18
Scale Max	63	63	49	49	42	126
Mean	41.52	50.20	26.52	31.67	34.66	81.58
SD	10.36	7.86	11.44	8.72	5.04	20.62

Table 5.3: Scores on Gold-MSI questionnaire, where numbers in brackets denote percentile of score. Lower part of the table: Norms based on 147,633 participants from Müllensiefen et al. (2014).

The results of the Gold-Genre task are scored as total correct pairs (of two) per participant. Every possible pair in every classification group is counted. This way, a total of 24 pairs is possible. Twin A scored 4 out of 24 possible pairs (15th percentile) and twin C scored 7 pairs (44th–53rd percentile). A's performance was thus again rather low (comparable to her performance on the questionnaire), while C performed better than her twin.

The Gold-Melody task can be scored using either accuracy or *d'*. Twin A obtained an accuracy score of 0.69 (36th–40th percentile) and a *d'* score of 0.93 (26th–30th percentile). Twin C performed well with an accuracy score of 0.85 (76th–80th percentile) and a *d'* score of 2.58 (76th–80th percentile).

The Gold-BAT task can again be scored using either accuracy or *d′*. Twin A obtained an accuracy score of 0.47 (1st–5th percentile) and a *d′* score of 0.38 (16th–20th percentile). Twin C obtained an accuracy score of 0.53 (6th–10th percentile) and a *d′* score of 0.58 (21st–25th percentile), thus neither of them performed well on the beat alignment subtest. Corroborating results were obtained with the perceptual part of the Beat Alignment Test by Iversen & Patel (2008), with which we additionally tested the twins. Since part of this test is identical to Gold-BAT and no scores of amusics are available for comparison, we refrain from reporting the detailed results.

3.2. Pitch Perception and Pitch Memory Abilities

For the pitch detection task, both A and C had thresholds of 0.14 semitones, which is comparable to the value Williamson & Stewart (2010) found for their control group. The threshold for their amusic group was slightly, however not significantly, higher, cf. first row in Table 5.4.

With respect to pitch direction discrimination, twin A reached a threshold of 0.55 semitones, which is considerably higher than her sister's: Twin C had a threshold of 0.15 semitones, again comparable to the values of Williamson & Stewart (2010) and also to those by Schaal et al. (2015), cf. second row in Table 5.4.

In the memory span task, both twins had comparatively low pitch spans of 3.5 and 3.3 tones respectively, which is comparable to the amusics' results in the previous studies, cf. first row in Table 5. Interestingly, there was a substantial difference in the performance of the two twins for the visual memory task: A had a visual memory span of only 2.2 letters, while C had a visual memory span of 8.8 letters. This is in contrast to Schaal et al.'s (2015) and Williamson et al.'s (2010) data, in which the amusics did not differ significantly from the controls in their visual memory span, cf. second row in Table 5.5.

Task	Twins		Williamson & Stewart 2010		Schaal et al. 2015	
	A	C	Amusic N=14	Control N=14	Amusic N=8	Control N=8
Pitch detection threshold	0.14	0.14	0.28 (±0.18)	0.14 (±0.03)	0.44 (±0.22)	0.36 (±0.22)
Pitch direction threshold	0.55	0.15	0.95 (±0.50)	0.17 (±0.05)	0.89 (±0.31)	0.19 (±0.03)

Table 5.4: Results of pitch detection and pitch direction task and results from Williamson & Stewart (2010) and Schaal et al. (2015) for comparison, all in semitones. Values in brackets indicate 95% confidence interval.

Task	Twins		Schaal et al. 2015		Williamson & Stewart 2010	
	A	C	Amusic N=8	Control N=8	Amusic N=14	Control N=14
Pitch Span	3.3	3.5	3.94	5.4	4.13	6.80
Visual Span	2.2	8.8	5.92	6.82	6.88	7.57

Table 5.5: Results of auditory memory span task (in tones) and visual memory span task (in letters), and for comparison the results from Williamson & Stewart (2010) and Schaal et al. (2015). Schaal et al. employed the same visual stimuli (Devanagari letters) as the present study, while Williamson et al. used digits instead.

3.3. Language Perception Abilities

3.3.1. Intonation Perception

The performance measures hit rate, percentage correct and *d'* for the intonation perception task are given in Table 5.6. Both twins differ on all three, and exhibit comparable performance to Hamann et al.'s (2012) cohort of amusics and controls respectively. For further analysis of the different semitone interval steps, only *d'* values were calculated.

Performance measure	Twins		Hamann et al. 2012	
	A	C	Amusic N=7	Control N=35
Hit rate	0.26	0.67	0.48 (±0.15)	0.78 (±0.46)
Percentage correct	62.05	80.80	68.58 (±5.75)	81.86 (±20.82)
d′	1.45	2.05	1.28 (±0.27)	2.00 (±1.17)

Table 5.6: Results of intonation task cumulated across all data (speech and sine analog together) and Hamann et al.'s (2012) for comparison. Values in brackets indicate 95% confidence interval.

Table 5.7 displays these *d′* values for the speech stimuli only and for the combination of both speech and sine analog stimuli. The latter is comparable to the results by Hamann et al. given in the last column of the table. As Table 5.7 shows, the *d′* values by A and hence her discriminatory ability were consistently lower at all interval sizes in comparison to her twin C.

Interval	Speech only		All data		Hamann et al. 2012	
	Twin A	Twin C	Twin A	Twin C	Amusics	Controls
1	0.00	0.65	-0.23	0.46	0.49	0.66
2	0.00	1.48	-0.23	1.12	0.51	1.42
3	0.00	2.48	-0.23	2.10	0.96	1.88
4	2.64	2.48	1.94	2.29	1.29	2.12
5	2.01	2.95	1.61	2.50	1.44	2.53
6	2.33	4.13	2.10	3.94	1.72	2.63
7	3.00	4.13	2.26	3.94	1.88	2.90

Table 5.7: *d′* values of C and A for each interval size for the speech stimuli and for both speech and sine analog stimuli together, and for comparison *d′* values from Hamann et al. (2012) across all of their data (speech, sine analog and pulse train analog).

The perceptual thresholds for A and C were calculated (for speech and sine analogs together) as elaborated in Hamann et al. (2012), resulting in a perceptual threshold of 5.39 semitones for A and 1.91 semitones for C. The thresholds for the speech stimuli, only, were 6.01 semitones for A and 2.25 semitones for C. This indicates that A's perception only reaches above chance performance at a difference of more than 5 semitones and is impaired in comparison to her sister's. These values are comparable to (though higher than) Hamann et al.'s findings of 3.80 semitones for their amusic group and 1.67 semitones for their control group.

3.3.2. Vowel Perception

For the vowel perception task, twin A had lower *d′* values on average and thus a lower discriminatory ability (Mean = 1.56, SE = 0.24) than twin C (Mean = 2.77, SE = 0.25), see Figure 5.3. This difference was significant at $t(94) = 3.46$, $p = 0.001$. Further analysis revealed that both twins have more difficulty with the shorter inter-stimulus interval (ISI) of 0.2 s, as shown by lower discriminatory abilities (*d′* values) in Figure 5.4.

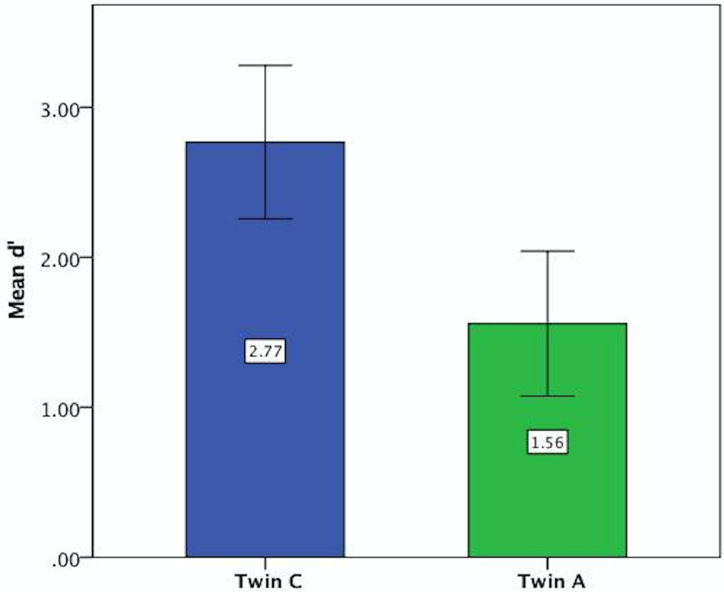


Figure 5.3: Mean *d′* values on vowel task showing lower discriminatory ability of Twin A. Error bars: 95% CI.

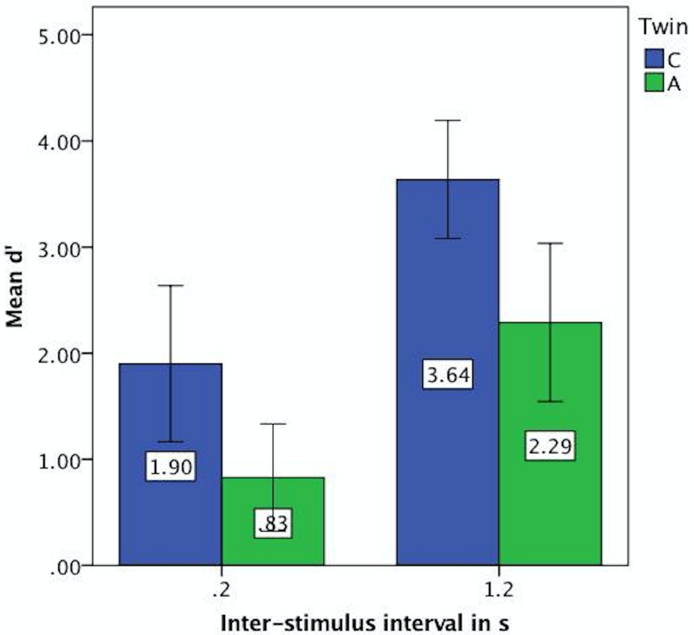


Figure 5.4: Mean *d'* values on vowel task split by inter-stimulus interval showing lower discriminatory ability for both twins at 0.2 s. Error bars: 95% CI.

Next, based on the percentage correct, we calculated a discrimination curve per continuum per participant, further split by the inter-stimulus interval. The resulting discrimination curves are visible in Figure 5.5. The steep discrimination curves for the longer inter-stimulus interval of 1.2 s, displayed in the right panel, show a clear categorization boundary. C's boundary is located between the 3rd and the 5th stimulus for all continua, which is to be expected. A's boundary is located between the 3rd and the 5th stimulus for the durational continua and between the 2nd and the 4th for the spectral continua. For the short inter-stimulus-interval of 0.2 s, the curves are not as steep in the boundary region. Especially A does not exhibit a steep slope for the durational continua.

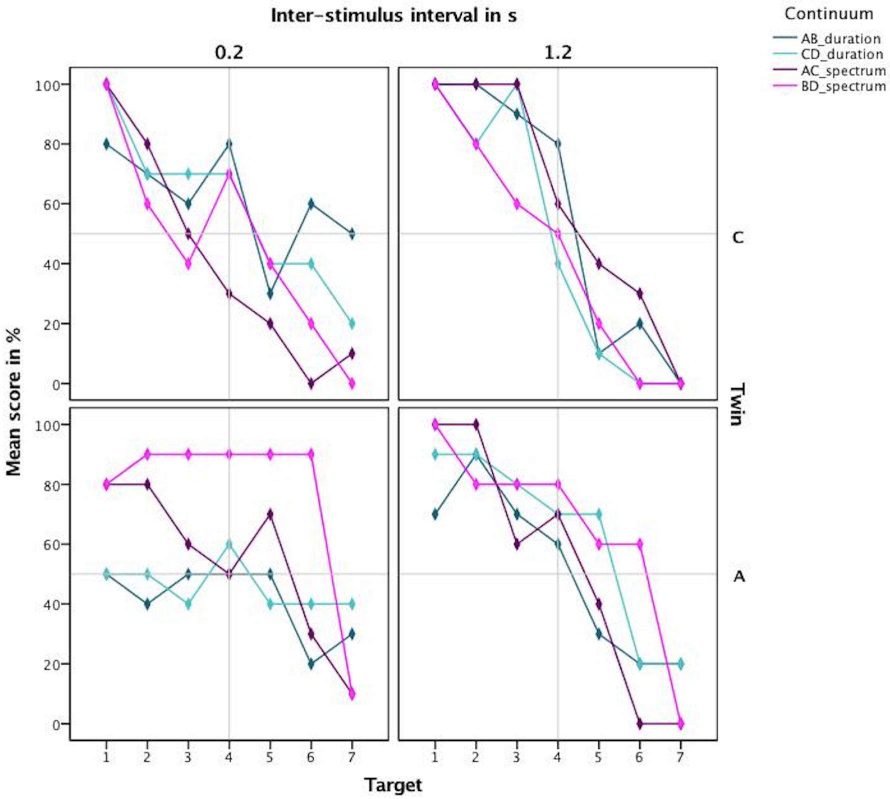


Figure 5.5. Discrimination curves based on mean scores in percent for both twins highlighting the different categorization boundary for the amusic twin for spectral cues.

3.4. Spatial Abilities

On the Object Perspective Taking Test, Twin A had a higher degree of deviation than twin C, namely 33.25º compared to 24.58º. The scores of both twins are above the sample mean of 24.53 by Hegarty & Waller (2004), and do not differ from each other significantly, as shown by an independent samples *t*-test $t(22) = -0.577$, $p > 0.05$.

On the Santa Barbara Solids Test, A achieved only a score of 20% correct answers, while C scored 83% correct answers and is above the sample mean of 68% provided by Cohen & Hegarty (2012). With respect to mistakes that show a failure in change of perspective, A made 50% and C 7% of such egocentric mistakes; the sample mean by Cohen & Hegarty is 19%.

There are no cut-off scores given for either of the two tasks. However, if two standard deviations are subtracted from the mean and this is taken as a cut-off score, than A's performance is still clearly an outlier on the Santa Barbara Solids Test, see Table 5.8.

Test	**Twin A**	**Twin C**	**Hegarty & Waller 2004 N=62**	**Cohen & Hegarty 2012 N=223**
Object Perspective Taking Test mean degree of deviation	33.25	24.58	24.53 S.D. 14.29	
Santa Barbara Solids Test score in absolute numbers out of 30	6 (15)	25 (2)		
Santa Barbara Solids Test score in percent	20% (50%)	83% (7%)		68% (19%) S.D. 23% (11%)

Table 5.8: Results of twins on Object Perspective Taking Test (Hegarty & Waller, 2004) and the Santa Barbara Solids Test (Cohen & Hegarty, 2012) with norm values. Value in brackets on the SBST indicates egocentric transformation mistakes.

4. Discussion

In this twin case study, we tested a dizygotic twin pair with one amusic twin and one non-amusic twin. Both twins had normal hearing and above average intellectual abilities, the latter also reflecting their higher than average education, both being graduate students at the time of testing (Asendorpf, 2009). Musical exposure and education of the twins was as comparable as it can be for two individuals, we can therefore conclude that congenital amusia is not due to differences in musical education or to insufficient exposure to music in childhood or adolescences as previously discussed by e.g. Peretz 2001.

A comprehensive overview of the twins' abilities as tested in this study is given in Table 5.9.

Ability	**Task**	**Twin A (amusic)**	**Twin C (non-amusic)**
Musical	Questionnaire	impaired	(✓)
	Gold-Genre	impaired	✓
	Gold-Melody	impaired	✓
	Gold-BAT	impaired	impaired
Pitch perception and memory	Detection	✓	✓
	Direction	impaired	✓
	Pitch Span	impaired	impaired
	Visual Span	impaired	✓
Language perception	Intonation	impaired	✓
	Vowel	impaired	✓
Spatial	Orientation	✓	✓
	Visualization	impaired	✓

Table 5.9: Overview of assessed abilities and results per twin.

Besides the MBEA (Peretz. et al. 2003), which clearly diagnosed one twin as amusic and the other as non-amusic, we employed the Goldsmith Musical Sophistication Index (Muellensiefen et al. 2014) to test their musical abilities further. Its self-report questionnaire reflects both twins' comparably low musical education (four years), but still clearly differentiates the twins, with the amusic twin always scoring in the lowest percentile. A slight exception is the factor of Active Engagement, where both twins score in the lowest two percentiles. Clear differences for the twins also emerge on the Gold-Genre and Gold-Melody subtests, with the non-amusic twin outperforming the amusic one. Only on the Gold-BAT subtest is their performance very similar and in a rather low range. This finding is not in line with the performance of the non-amusic twin on the MBEA Rhythm subtest, which was very high (her highest score on any of the subtests), while the amusics' score was very low. A larger study of the Gold-MSI with amusics should be conducted in the future to see whether the pattern shown by the amusic in this study holds for a larger group of amusics, i.e. whether the Gold-MSI can be used to reliably differentiate amusics from non-amusics. In addition, the MBEA is very repetitive and tedious to complete for amusics, while the Gold-MSI offers different tasks and has a questionnaire already included. So, future directions might be to use the Gold-MSI in addition to the MBEA or possibly even as a replacement, since the MBEA has an imbalance in pitch and rhythm-based subtests, as was already pointed out by Pfeifer & Hamann (2015).

The finding that both twins have a comparable low pitch detection threshold of 0.135 tones (indicating no impairment), while their pitch direction threshold differs, is in line with previous findings on amusics (Williamson & Stewart, 2010), and indicates that their auditory processing is unimpaired but that congenital amusia has an impact on the perception of changes in pitch direction. It is surprising that both twins exhibit a low pitch memory span in comparison to normal controls (Schaal et al., 2015), which might be interpreted as an indication for a certain hereditariness of pitch memory, as has been proposed for pitch processing (Drayna et al., 2001). Mosing et al. (2014) report a positive association for their large twin cohort between the different auditory tasks for the twin pairs. They find that this is mostly due to shared genes and to a smaller degree to shared environmental factors affecting musical abilities. This leads back to a nature versus nurture debate and ties into the question of the genetic underpinnings of congenital amusia. The dizygotic twin pair share 50% of their genes and we can assume that congenital amusia – since it is only present in one twin – is somehow encoded in the 50% of non-shared genes. What is puzzling in the present case, however, is that both twins exhibit pitch memory impairments. These could either be due to their 50% of shared genes or to their shared environment. In the future, gene sequencing of congenital amusia is required to unravel the underpinnings of this disorder and to further understand the genetics of musical abilities and general auditory processing. The dizygotic twin pair discussed in this article and a further amusic monozygotic twin pair that we have just identified seem to be a promising starting point for a genetic analysis.

While the everyday communication of the amusic twin seems to be unimpaired and her score on the verbal subscale of the Hamburger Wechsler Adult Intelligence Scale is high (124 IQ points compared to 111 by the non-amusic twin), her intonation perception and vowel perception are impaired in comparison to her sister, and she shows overall lower discriminatory abilities. This was to be expected based on previous studies on language perception by amusics (e.g. Liu et al., 2010; Hamann et al., 2012). Interestingly, the stimuli without linguistic information on the intonation perception task resulted in better performance for both twins than real speech stimuli. Future studies with amusics and controls need to test whether the presence of linguistic in addition to tonal information does not enhance pitch perception. On the vowel perception task, A exhibited a different categorical boundary for spectral cues than her twin. We are currently conducting a study on vowel perception with a larger pool of amusics and controls in order to investigate the pattern exhibited by the twins.

Lastly and most surprisingly, the twins also performed differently on one of the spatial tasks, with the non-amusic twin (83% correct) outperforming the amusic twin (20% correct). Taken together, the results indicate that the amusic twin can perform egocentric spatial transformations, as shown by the Object Perspective Taking Test, but struggles with object-based spatial transformations that were required in the Santa Barbara Solids Test. Her sister had no difficulties with the latter. This shows that at least this one amusic has impaired spatial visualization abilities with intact spatial orientation abilities. Our finding contrasts with that by Tillmann et al. (2010) who assume spatial abilities by amusics to be unimpaired based on their test, but are in line with Douglas and Bilkey's (2007) finding, the self-reports given in Peretz & Vuvan (2017) and the longer reaction time latencies found by Williamson et al. (2011) for a subgroup of amusics. These indications for a very specialized impairment warrant further scrutiny of amusics' spatial abilities and a fractionating of their skills in this regard.

5. Conclusion

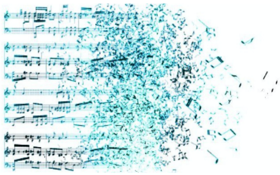
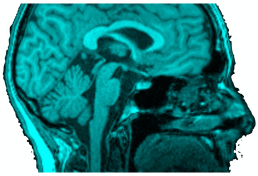
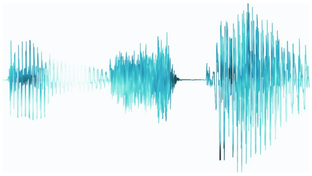
This study was the first to employ the Goldsmith Musical Sophistication Index to test the differences between an amusic and a non-amusic participant. All in all, the Gold-MSI seems to be able – at least in this very limited sample – to differentiate between non-amusic and amusic participants. In the future, a larger sample of amusics should be tested with it to assess whether this holds true for a larger group. If this is the case, the Gold-MSI could be used to supplement or possibly replace the MBEA in the diagnosis of congenital amusia in the future.

We also showed that the question of a spatial processing deficit in amusia needs to be revisited and more research is needed in that area. Most notably, separate tests should be employed for egocentric and object-based spatial transformations to be able to differentiate between the two, as only the latter turned out to be impaired in our amusic twin.

This twin case study highlights that congenital amusia is not due to insufficient exposure to music in childhood. The exposure to music of the twin pair was as comparable as it can be for two individuals. Yet, one twin has amusia, while the other does not. In addition to the expected differences in melodic and language perception abilities, we found that both twins exhibit a comparably low pitch memory span and low beat perception abilities. This raises the question of nature versus nurture, i.e. whether their shared genes or their shared environment and low musical education is responsible for the shared deviant performance. This in turn gives rise to the question of hereditariness of congenital amusia and calls for a genetic analysis of affected individuals. To prove that genetic causes play a role in congenital amusia, a large-scale genetic analysis of amusics and their unaffected relatives is necessary. From such a study, we could learn more about how amusia can develop and could identify which genes contribute to higher cognitive functions of auditory perception.

6. Acknowledgments

We would like to thank A and C for their endless patience in fulfilling all tasks that we gave them.



6. Conclusion

In this thesis we originally set out to investigate speech perception impairments in congenital amusia: Whether there were any, and which parts of speech and which linguistic cues might be affected. Before we were able to undertake our planned studies, however, we realized that we needed more reliable standards for the diagnosis of congenital amusia. Out of this realization arose the large-scale study in Chapter 2 (published as Pfeifer & Hamann, 2015), in which we tested a large cohort of students with the *Montreal Battery of Evaluation of Amusia* (MBEA; Peretz, Champod, & Hyde, 2003). We established that the Signal Detection Theory measure *d'* should be used for scoring the MBEA, that non-parametric statistics ought to be used, and that a composite score calculated across the different subtests is not useful as it does not capture the different subtypes of amusics that we found. We also showed that the prevalence of amusics depended strongly on the used cut-off score and number of subtests.

The scoring procedure with Signal Detection Theory, which we proposed, has now been widely adopted by other research groups (e.g. Lu, Ho, Sun, Johnson, & Thompson, 2016; Zhou, Liu, Jing, & Jiang, 2017; Sun et al., 2018; Toledo-Fernández, Villalobos-Gallegos, García-Gómez, & Salvador-Cruz, 2018; Corrow et al., 2019; Zhou, Liu, Jiang, & Jiang, 2019; Zhou, Liu, Jiang, Jiang, & Jiang, 2019). We also advocated the use of questionnaires. Furthermore, we showed that the MBEA is not suitable for a web-based diagnosis of congenital amusia, despite the recent rise in web-based testing.

Vuvan et al. (2018) have since reacted to our critique and have in response published the *Montreal Protocol for the Identification of Amusia* (MPIA). In their article, the authors argue that the MBEA should merely be seen as part of a larger screening procedure employed by them and that so far they had omitted to publish the other parts of the screening procedure that they employ (Vuvan et al., 2018). However, in my opinion the MPIA does not resolve the issues concerning the MBEA that we raised in Chapter 2. The MBEA is still the main part of the MPIA, and the empirical and statistical problems that we pointed out were not tackled in the MPIA; Vuvan et al. (2018) merely interpret them as strengths instead. They are also not resolved by reframing the MBEA as a screening instead of a diagnostic test. The issues mentioned by Vuvan et al. should therefore be discussed in detail in a future study to clarify the usefulness of the MPIA.

In the long run, a more reliable screening tool ought to be designed. Ideally, this tool should not be solely based on music, as amusia has time and again, including in this thesis, been shown to also affect language in very specific ways. Therefore, a new possible test or testing protocol should either include music as well as speech components or use more general pitch- and duration-based tests. We explored the usage of the *Goldsmith Musical Sophistication Index* (GoldMSI, Müllensiefen, Gingras, Musil, & Stewart, 2014) and its various components as a possible screening tool in Chapter 5. The results seem promising. However, validation with a large group of amusics would be needed and the GoldMSI would also only be able to serve as the music component of a larger screening protocol.

After having established at least somewhat more reliable standards for the diagnosis of amusia, we tested amusics' vowel perception in Chapter 3 and their word stress perception in Chapter 4. In both chapters an electrophysiological study was preceded by a behavioral study.

In Chapter 3 we investigated formant frequencies and duration as cues in vowel perception, finding first of all that amusics performed behaviorally worse in discriminating vowel pairs, showing perception deficits both for vowel quality and vowel quantity. In addition, we found that amusics do exhibit an MMN, but their MMN is significantly reduced in comparison to controls.

In Chapter 4 we investigated pitch and duration differences as cues for word stress perception and found somewhat different results concerning cue usage in comparison to Chapter 3: In the behavioral task, amusics did not struggle with pitch cues and even relied more on durational cues than controls, which could point to possible compensation strategies. The findings concerning the MMN were similar to the findings in Chapter 3 however. Amusics displayed an MMN but it was significantly reduced in comparison to controls.

While we wanted to answer questions about speech perception and congenital amusia in this thesis, we arrived at the findings very briefly summarized above. On the one hand, we also faced several constraints or limitations in our research that led to some aspects still being open and on the other hand, many new questions and possible research topics arose. These will be discussed in the following paragraphs.

Most broadly, in our studies in Chapters 3 and 4 we investigated vowel and word stress perception only. There are many other aspects of language that we did not even touch upon but that are also worth investigating in congenital amusia. We, for example, never investigated consonants, which other researchers have started doing. An interesting starting point here would be the voice onset time of obstruents, which has recently been investigated by Zhang, Shao, & Huang (2017) and Jasmin, Dick, Holt, & Tierney (2019). Due to experimental design constraints, we had to limit our investigation of vowels to formant and durational cues and to pitch and duration for word stress. Not only do more cues remain untested, we also only tested one group of vowels and only one word stress pair. In addition, we only used synthesized, isolated vowels. Our study might also be interesting to replicate with natural, edited vowels and with vowels embedded in syllables.

One question that arose from our vowel study is whether all amusics would have problems with durational cues: In our studies, we only included amusics who had a pitch and a rhythm perception deficit at the same time, in an attempt to keep the groups as comparable as possible. The question remains, whether amusics who only have a pitch perception/memory deficit would also have difficulties with durational cues or whether this is limited to amusics with rhythmic disabilities. Also it is interesting that amusics seemed to rely more heavily on durational cues in our word stress study. This might point to a compensatory strategy that is, however, not repeated in our vowel perception study. These findings need further investigation to assess whether they can be somehow consolidated. As compensational strategies have never been investigated in detail in congenital amusia, this is also an area that requires further attention.

The underlying electrophysiology of speech processing in congenital amusia is another area in which we provided answers as well as raised further questions. While we managed to show that amusics at least exhibited an MMN as a marker of early auditory change detection, it was, however, always significantly reduced in comparison to controls. One thing to keep in mind is our still relatively small sample size for EEG studies. This reduction could simply be a sampling bias. Unfortunately, it is very difficult to include more amusics, as amusic testing cohorts are generally rather small. Due to various exclusion criteria for EEG studies, the number of participants has been limited even further. In comparison to other EEG studies with amusics, our sample size was normal to large. The specificity of this reduction is also still rather unclear and should be investigated further.

The first question that arose in this area is whether a phonological MMN would also be affected by amusia. In both our studies, we focused on an auditory MMN, as that seemed to be the necessary starting point to us. However, one of the first follow-ups one ought to consider is looking at the phonological MMN calculated across blocks instead of within one block.

Following this, and based on musical EEG studies, the next thing to investigate are further EEG components, such as the N400 in relation to cross-modal priming (e.g. Zhou et al., 2017) or the P300 (e.g. Peretz, Brattico, & Tervaniemi, 2005; Braun et al., 2008;) or P600 (e.g. Jiang et al., 2012a). An N400 cross-modal priming study has been carried out by us and is currently being analyzed (Pfeifer & Hamann in preparation).

Another very promising research field would be a combination of EEG together with non-invasive electrical brain stimulation, i.e. tACS or tDCS. Schaal, Pfeifer, Krause, & Pollok (2015) used tACS successfully to positively influence amusics' memory for pitch sequences. It remains to be seen whether these effects can also be found for speech sounds and whether brain stimulation might produce beneficial effects for other aspects than just memory. Other brain areas and other frequency ranges as employed by Schaal et al. (2015) should be investigated. Studies using tACS and EEG concurrently could explore the different neural oscillation and their loci in congenital amusia. This approach could in turn lead to a better understanding of the underlying deficits of congenital amusia and more generally to a better understanding of auditory processing and linguistic processing.

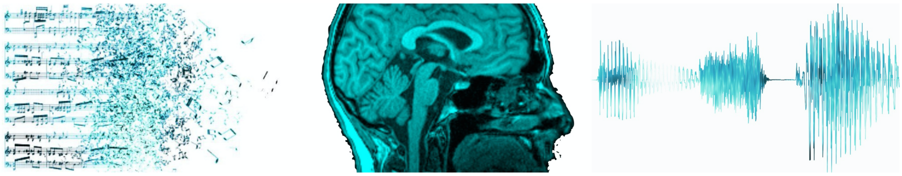
Again more broadly, we have only investigated speech perception. Another area that should still be investigated is speech production in congenital amusia. The different parameters that have been instigated by us in amusics' perception would now, logically, need to be examined in their production. Amusics seem not to have any production deficits on the surface, as their everyday communication is mostly unremarkable and unproblematic. We therefore either face a puzzling dissociation between cues that are not properly perceived but correctly produced, or amusics do have a slightly deviating production that accurate measurements in the laboratory might reveal.

In the last study in Chapter 5, we assessed the intellectual, musical and language abilities of a dizygotic female twin pair with one amusic and one non-amusic twin. We have shown that despite the same upbringing, education and exposure to music, amusia emerged in only one of the twins.

Thereby, we showed that amusia is not caused by a lack of exposure to music or musical education.
We also had two surprising findings. One was that both twins displayed a low pitch memory span. This could be interpreted as an indication for a certain hereditariness of pitch memory, as has been proposed for pitch processing (Drayna, Manichaikul, de Lange, Snieder, & Spector, 2001). This in turn ties into the question of the genetic underpinnings of congenital amusia and a nature versus nurture debate. A thorough genetic analysis of congenital amusia has not yet been carried out and so far, only anecdotal evidence has been provided (Peretz, Cummings, & Dube, 2007): Peretz et al. drew up 9 family trees consisting of a total of 380 individuals; 22 of which were amusics. However, they only behaviorally assessed 67 individuals. The musicality of the other 313 individuals was only reported via hearsay of their relatives. No genetic material was collected or analyzed. Therefore this evidence, while it points in a promising direction, can only be regarded as anecdotal. Furthermore, while Peretz and colleagues have collected genetic material of amusics and have performed whole exome sequencing, this data is not published or otherwise publicly available and is only referred to in conference talks by Isabelle Peretz. Hopefully this data will be published in the future. Meanwhile, we have performed a whole-genome analysis of a first amusic sample (N=9). This sample consist of a family with 4 generations and one amusic individual per generation (Pfeifer & Lüthy in preparation).

The other interesting finding was that the twins also performed differently on one of the spatial tasks, with the non-amusic twin outperforming the amusic twin, opening up a debate about different aspects of spatial perception in congenital amusia again. Concerning this question, we have also already collected and analyzed data.

To conclude, a lot of research about congenital amusia has been carried out in the last couple of years, of which the articles in this thesis are only a small part. We hope that the findings in this thesis contribute to disentangling at least some aspects of perception in congenital amusia. This disorder still offers many interesting, unsolved questions, some of which have been outlined in this chapter. Generally, it is hoped that congenital amusia and its cognitive, neural and genetic underpinnings are one day completely revealed and understood. This in turn could lead to a better understanding of what is shared between music and language processing, the neural and cognitive underpinnings of auditory perception in general, and - very generally - provide information about neuroplasticity and the genetic pathways involved in auditory/language perception. It is, therefore, important to continue research into the scientific mysteries that congenital amusia still offers.



7. References

Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research, 812*(1), 23–37.

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception memory deficits in congenital amusia. *Scientific Reports, 6*, 18861.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.–E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain, 136*(5), 1639–1661.

Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research, 1537*, 224–232.

Anderson, S., Himonides, E., Wise, K., Welch, G., & Stewart, L. (2012). Is there potential for learning in amusia? A study of the effect of singing intervention in congenital amusia. *Annals of the New York Academy of Sciences, 1252,* 345–353.

Asendorpf, J. (2009). *Persönlichkeitspsychologie* (4th ed.). Heidelberg: Springer.

Auditory (2007). *Online listening tests and psychoacoustic experiments with large N*. Discussion on Auditory mailing list, see www.auditory.org/mhonarc/2007/threads.html#00531.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia - A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238–251.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995) *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology, 37*(1), 13–20.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology, 55,* 803–832.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.2.25). Retrieved from www.praat.org/

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.15). Retrieved from www.praat.org/

Bones, O., & Wong, P. C. M. (2017). Congenital amusics use a secondary pitch mechanism to identify lexical tones. *Neuropsychologia, 104*, 48–53.

Braun, A., McArdle, J., Jones, J. L., Nechaev, V., Zalewski, C., Brewer, C., & Drayna, D. (2008). Tune deafness: Processing melodic errors outside of conscious awareness as reflected by components of the auditory ERP. *PLoS ONE, 3*(6), 1–6.

Chen, J., Kumar, S., Wiliamson, V., Scholz, J., Grifiths, T., & Stewart, L. (2014). Detection of the arcuate fasciculus in congenital amusia is dependent on tractography algorithm. *Proceedings of the 20th Annual Meeting of the Organization for Human Brain Mapping, 6*, 1–11.

Chen, J., & Yuan, J. (2016). The neural causes of congenital amusia. *The Journal of Neuroscience*, *36*(30), 7803–7804.

Chládková, K., Escudero, P., & Lipski, S. C. (2013). Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. *Brain and Language, 126*(3), 243–252.

Cohen, C. A., & Hegarty, M. (2012). Inferring cross sections of 3D objects: A new spatial thinking test. *Learning and Individual Differences, 22*(6), 868–874.

Corrow, S. L., Stubbs, J. L., Schlaug, G., Buss, S., Paquette, S., Duchaine, B., & Barton, J. J. S. (2019). Perception of musical pitch in developmental prosopagnosia. *Neuropsychologia, 124*, 87–97.

Cousineau, M., Mcdermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences, 109,* 19858–19863.

Cousineau, M., Oxenham, A. J., & Peretz, I. (2015). Congenital amusia: A cognitive disorder limited to resolved harmonics and with no peripheral basis. *Neuropsychologia, 66*, 293–301.

Cowan, G. M. (1948). Mazateco whistle speech. *Language, 24*, 280–286.

Cox, J. (1948). Tone deafness. *Music Educator's Journal, 34*(4), 62–63.

Cuddy, L. L., Balkwill, L.-L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare - A study of "Tone Deafness" among university students. *Annals of the New York Academy of Sciences, 1060***,** 311–324.

Cutini, S., Szűcs, D., Mead, N., Huss, M., & Goswami, U., (2016). Atypical right hemisphere response to slow temporal modulations in children with developmental dyslexia, *NeuroImage, 143*, 40–49.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201–220.

Dalla Bella, S., Berkowska, M., & Sowiński, J., (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology, 2*(164).

Dandurand, F., Shultz, T., & Onishi, K. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods, 40,* 428–434.

De Bree, E., Wijnen, F., & Zonneveld, W. (2006). Word stress production in three-year-old children at risk of dyslexia. *Journal of Research in Reading*, *29*, 304–317.

de Jonge, M. J. I., & Boersma, P., (2015). French high-mid vowels are underspecified for height. *Proceedings of the 18th International Congress of Phonetic Sciences*, Paper 948.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). *The Visual Patterns Test: A new test of short-term visual recall*. Suffolk, England: Thames Valley Test Company.

Dogil, G. (1995). The phonetic manifestation of stress. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Universität Stuttgart), 2*, 3–51.

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience, 10*, 915–921.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science, 291*, 1969–1971.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language, 36*(3), 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition, 106*(2), 682–706.

Eggeling, J. (1882). *The Satapatha-Brahmana* (J. Eggeling, Trans.). Oxford: Claredon Press.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551–585.

Fiedler, D., & Müllensiefen, D. (2015). Validation of the Gold-MSI questionnaire to measure musical sophistication of German students at secondary education schools. *Proceedings of Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*.

Fikkert P., & Penner Z. (1998). Stagnation in prosodic development of language-disordered children. *Proceedings of the Boston University Conference on Language Acquisition*, 201–212.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T., D. (2004). Characterization of deficits in pitch perception underlying “tone deafness”. *Brain, 127*, 801–810.

Foxton, J. M., Nandy, R. K., & Griffiths, T. D. (2006). Rhythm deficits in ‘tone deafness’. *Brain and Cognition, 62*(1), 24–29.

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including Linear Mixed Models. *Frontiers in Neuroscience, 12*(48).

Frost, R. (1969). *The Poetry of Robert Frost: The Collected Poems, Complete and Unabridged*. New York: Henry Holt.

Gallon, N., Harris, J., & Van der Lely, H. (2007). Non-word repetition: An investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics & Phonetics*, *21*, 435–455.

Geschwind, N. (1984). The brain of a learning-disabled individual. *Annals of Dyslexia, 34*, 319–327.

Giegerich, H. (1985). *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press.

Gingras, B., Honing, H., Peretz, I., Trainor, L. J., & Fisher, S. E. (2015). Defining the biological bases of individual differences in musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664).

Gordon, E. E. (1965). *Muscial Aptitude Profile.* Boston: Houghton Mifflin.

Gosselin, N., Jolicœur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences, 1169*(1), 270–272.

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language, 69*, 1–17.

Grant, A. (1878). Note-deafness. *Mind, 10*, 157–167.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Haake, C., Kob, M., Willmes, K. & Domahs, F. (2013). Word stress processing in specific language impairment: Auditory or representational deficits? *Clinical Linguistics & Phonetics*, *27*(8), 594–615.

Hamann, S., Exter, M., Pfeifer, J., & Krause-Burmester, M. (2012). Perceiving differences in linguistic and non–linguistic pitch: A pilot study with German congenital amusics. *Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM)*, eds. Cambouropoulos, E., Tsougras, C., Mavromatis P., & Pastiadis, K., 398–405.

Harmon, R. (2003). From themistocles to philomathes: Amousos and amousia in antiquity and the early modern period. *International Journal of the Classical Tradition, 9*(3), 351–390.

Hay, R. A., Roach, B. J., Srihari, V. H., Woods, S. W., Ford, J. M., & Mathalon, D.H. (2015). Equivalent mismatch negativity deficits across deviant types in early illness schizophrenia-spectrum patients. *Biological psychology*, *105*, 130–137.

Hayward, K. (2000). *Experimental Phonetics*. Harlow: Pearson Education Limited.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective–taking spatial abilities. *Intelligence, 32*(2), 175–191.

Henry, M. J., & McAuley, J. D. (2010). On the prevalence of congenital amusia. *Music Perception: An Interdisciplinary Journal, 27*, 413–418.

Henry, M. J., & McAuley, J. D. (2013). Failure to apply Signal Detection Theory to the Montreal Battery of Evaluation of Amusia may misdiagnose amusia. *Music Perception: An Interdisciplinary Journal, 30*(5), 480–496.

Honing, H., & Ladinig, O. (2008). The potential of the internet for music perception research: A comment on lab-based versus web-based studies. *Empirical Musicology Review, 3,* 4–7.

Huang, W. T., Liu, C., Dong, Q., & Nan, Y., (2015). Categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Frontiers in Psychology*, *6*(829).

Huang, X, Zhang, C., Shi, F., Yan, N., & Wang, L. (2016) Impaired vowel discrimination in Mandarin-speaking congenital amusics. *Proceedings of the 5th International Sumposium on Tonal Aspects of Language*.

Hutchins, S., Gosselin, N., & Peretz, I. (2010a). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology, 1*, 1–8.

Hutchins, S., & Peretz, I. (2013). Vocal pitch shift in congenital amusia (pitch deafness). *Brain and Language, 125***,** 106–117.

Hutchins, S., Zarate, J., Zatorre, R., & Peretz, I. (2010b). An acoustical study of vocal pitch matching in congenital amusia. *Journal of the Acoustical Society of America, 127,* 504–512.

Hyde, K., Lerch, J. P., Zatorre, R., Griffiths, T., D., & Evans, A. (2007). Cortical thickness in congenital amusia: When less is better than more. *The Journal of Neuroscience, 47*, 13028–13032.

Hyde, K., & Peretz, I. (2003). “Out-of-Pitch” but still “In-Time”. *Annals of the New York Academy of Sciences, 999,* 173–176.

Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356–360.

Hyde, K., Zatorre, R., Griffiths, T., D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain, 129*, 2562–2570.

Hyde, K., Zatorre, R., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex, 21*(2), 292–299.

Iversen, J. R., & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. *Proceedings of the 10th International Conference on Music Perception and Cognition*, 465–468.

Iverson, P. (2012). Measuring phonetic perception in adults, in *The Oxford Handbook of Laboratory Phonology*, eds. A. C. Cohn, C. Fougeron & M. Huffman. Oxford: Oxford University Press, 572–580.

Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2019). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*. Advance online publication.

Jessen, M. (1993). Stress conditions on vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory*, *8*, 1–27.

Jessen, M., Marasek, K., Schneider, K., & Classen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. *Proceedings of the International Congress of Phonetic Sciences, 4*, 428–431.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Chen, X., & Yang, Y. (2012a). Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension. *PLoS ONE, 7*, e41411.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia, 48*, 2630–2639.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012b). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition, 40,* 1109–1121.

Jiang, C., Lim, V. K., Wang, H., & Hamm, J. P. (2013). Difficulties with pitch discrimination influences pitch memory performance: Evidence from congenital amusia. *PLoS ONE, 8*, e79216.

Kalmus, H. (1948). Tune deafness and its inheritance. *Proceedings of the Eighth International Congress of Genetics*, Paper 605.

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics, 43*, 369–383.

Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication, 77*, 84–100.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology, 67*(2), 131–143.

Koerner, T. K., & Zhang, Y. (2017). Application of Linear Mixed-Effects Models in human neuroscience research: A comparison with pearson correlation in two auditory electrophysiology studies. *Brain sciences, 7*(3), 26.

Kohler, K. J. (1990). German. *Journal of the International Phonetic Association, 20*(1), 48–50.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research, in *Psychological Experiments on the Internet,* ed. M.H. Birnbaum. San Diego: Academic Press, 35–60.

Kujala, T., & Näätänen, R. (2001). The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neuroscience & Biobehavioral Reviews, 25*(6), 535–543.

Launay, J., Grube, M., & Stewart, L. (2014). Dysrhythmia: a specific congenital rhythm perception deficit. *Frontiers in Psychology, 5*, 1–18.

Lebrun, M.-A., Moreau, P., McNally-Gagnon, A., Mignault Goulet, G., & Peretz, I. (2012). Congenital amusia in childhood: A case study. *Cortex, 48*(6), 683–688.

Leconte, J. (1887). Sound-Blindness. *Science, 10*(250), 244–245.

Leong, V., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language, 64*, 59–73.

Lintfert, B. (2010). *Phonetic and phonological development of stress in German*. PhD dissertation, Universität Stuttgart, Stuttgart.

Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., & Stewart, L. (2013). Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing. *Attention, Perception, & Psychophysics, 75*, 1783–1798.

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012a). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PLoS ONE, 7*, e30374.

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia, 66*, 111–118.

Liu, F., Maggu, A. R., Lau, J. C. Y., & Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Frontiers in Human Neurosciences, 8*, 1–19.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain, 133*(6), 1682–1693.

Liu, F., Xu, Y., Patel, A. D., Francart, T., & Jiang, C. (2012b). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition, 79*, 209–215.

Lolli, S. L., Lewenstein, A. D., Basurto, J., Winnik, S., & Loui, P. (2015). Sound frequency affects speech emotion perception: Results from congenital amusia. *Frontiers in Psychology, 6*, 1–10.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience,* 29, 10215–10220.

Loui, P., Guenther, F.H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology, 18*, R331–R332.

Loui, P., Kroog, K., Zuk, J., Winner, E., & Schlaug, G. (2011). Relating pitch awareness to phonemic awareness in children: Implications for tone-deafness and dyslexia. *Frontiers in Psychology, 2*(111).

Loui, P., & Schlaug, G. (2009). Investigating musical disorders with diffusion tensor imaging. *Annals of the New York Academy of Sciences, 1169*(1), 121–125.

Loui, P., & Schlaug, G. (2012). Impaired learning of event frequencies in tone deafness. *Annals of the New York Academy of Sciences, 1252,* 354–360.

Lu, X., Ho, H. T., Sun, Y., Johnson, B. W., & Thompson, W. F. (2016). The influence of visual information on auditory processing in individuals with congenital amusia: An ERP study. *NeuroImage, 135*, 142–151.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: An auditory-motor feedback disorder? *Restorative Neurology and Neuroscience, 25*, 323–334.

Marin, M. M., Gingras, B., & Stewart, L. (2012). Perception of musical timbre in congenital amusia: Categorization, discrimination and short-term memory. *Neuropsychologia, 50*(3), 367–378.

Marin, M. M., Thompson, W. F., & Stewart, L. (2012). Emotion perception of dyads and triads in congenital amusia. *Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM).*

Mcdonald, C., & Stewart, L. (2007). Uses and functions of music in Congenital Amusia. *Music Perception, 25*, 345–356.

McGraw, K., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11,* 502–506.

McWhinney, S. R. (2018). *A comparison of statistical methods for relating individual differences to event-related potential components.* PhD dissertation, Dalhousie University Halifax, Nova Scotia.

Mehler, J. (1999). Editorial. *Cognition, 71*, 187–189.

Mignault Goulet, G., Moreau, P., Robitaille, N., & Peretz, I. (2012). Congenital Amusia persists in the developing brain after daily music listening. *PLoS ONE, 7*, e36860.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*, 272–277.

Moon, I. J., & Hong, S. H. (2014). What is temporal fine structure and why is it important? *Korean Journal of Audiology, 18*(1), 1–7.

Moreau, P., Jolicœur, P., & Peretz, I. (2009). Automatic brain responses to pitch changes in Congenital Amusia. *Annals of the New York Academy of Sciences, 1169*, 191–194.

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition, 81*, 337–344.

Mosing, M. A., Pedersen, N. L., Madison, G, & Ullén F. (2014). Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. *PLoS ONE, 9*(11), e113874.

Moulton, W. G. (1962). *The Sounds of English and German*. Chicago: University of Chicago Press.

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.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13*(2), 201–233.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1–21.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313–329.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590.

Näätänen, R., Sussmann, E.S., Salisbury D., & Shafer V. L. (2014). Mismatch negativity (MMN) as an index of cognitive dysfunction. *Brain Topography,* 27(4), 451–66.

Nan, Y., Huang, W.-t., Wang, W.-j., Liu, C., & Dong, Q. (2016). Subgroup differences in the lexical tone mismatch negativity (MMN) among Mandarin speakers with congenital amusia. *Biological Psychology, 113*, 59–67.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain*, *133*, 2635–2642.

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in Congenital Amusia. *Annals of the New York Academy of Sciences, 1169,* 490–493.

Omigie, D., Müllensiefen, D., & Stewart, L. (2012a). The experience of music in congenital amusia. *Music Perception: An Interdisciplinary Journal, 30,* 1–18.

Omigie, D., Pearce, M. T., & Stewart, L. (2012b). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia, 50,* 1483–1493.

Omigie, D., Pearce, M.T., Williamson, V.J., & Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia, 51,* 1749–1762.

Omigie, D., & Stewart, L. (2011). Preserved statistical learning of tonal and linguistic material in Congenital Amusia. *Frontiers in Psychology, 2*(109).

Partanen, E., Vainio, M., Kujala, T., & Huotilainen, M. (2011). Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles. *Psychophysiology, 48*(10), 1372–1380.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition, 59*, 310–313.

Patel, A. D., Wong, M., Foxton, J. M., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception, 25*, 357–368.

Pätzold, M., & Simpson, A. P. 1997. Acoustic analysis of German vowels in the Kiel corpus of read speech. *Arbeitsbreichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel, 32*, 215–247.

Peretz, I. (2001). Brain specialization for music: New evidence from congenital amusia. In: Biological foundations of music. *Annals of the New York Academy of Sciences, 930*, 153–165.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*, 185–191.

Peretz, I., Blood, A. J., Penhune, V., & Zarorre, R. (2001). Cortical deafness to dissonance. *Brain, 124*, 928–940.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: In tune, out of key, and unaware. *Brain, 132*, 1277–1286.

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in Congenital Amusia. *Annals of Neurology, 58,* 478–482.

Peretz, I., Champod, S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*, 58–75.

Peretz, I., Cummings, S., & Dube, M.-P. (2007). The genetics of congenital amusia (Tone Deafness): A family-aggregation study. *The American Journal of Human Genetics, 81*, 582–588.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., & Bouchard, B. (2008). On-line idenfication of congenital amusia. *Music Perception, 25*, 331–343.

Peretz, I., Saffran, J., Schön, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences, 1252*, 361–366.

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics, 25*(5), 625–630.

Pettinato, M., & Verhoeven, J. (2008). Production and perception of word stress in children and adolescents with Down syndrome. *Down Syndrome Research and Practice*.

Pfeifer, J., & Hamann, S. (2014). Congenital Amusia in linguistic and non-linguistic pitch perception - What behavior and reaction times reveal. *Proceedings of Speech Prosody 7*, eds. Campbell, N. Gibbon, D., & Hirst, D., 438–442.

Pfeifer, J., & Hamann, S. (2015). Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia. *Frontiers in Human Neurosciences, 9*, 1–15.

Pfeifer, J. & Hamann, S. (2015). Web-based testing of congenital amusia with the Montreal Battery of Evaluation of Amusia. *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, eds. Ginsborg, J., Lamont, A., Phillips, M., Bramley, S.

Pfeifer, J., & Hamann, S. (2018). The Nature and Nurture of Congenital Amusia: A Twin Case Study. *Frontiers in Behavioural Neurosciences* 12: 120.

Pfeifer, J., & Hamann, S. (in preparation). Dutch word stress perception by congenital amusics.

Pfeifer, J., & Lüthy, K. A. (in preparation). A whole genome analysis of an amusic family.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., & Peretz, I. (2013). Amusic does not mean unmusical: Beat perception and synchronization ability despite pitch deafness. *Cognitive Neuropsychology, 30*, 311–331.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia, 49,* 961–969.

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch Negativity: Different water in the same river. *Audiology and Neurotology, 5*(3–4), 111–139.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*(1), 221–231.

Provost, M. (2011). *The Prevalence of congenital amusia.* Master's thesis, Université de Montréal.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. See www.R-project.org/.

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions, in *Psychological Experiments on the Internet,* ed. M.H. Birnbaum, San Diego: Academic Press, 89–117.

Reips, U.-D. (2000). Standards for internet-based experimenting. *Experimental Psychology, 49*, 243–256.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20*(3), 403–452.

Schaal, N. K., Bauer, A.-K. R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe. *Musicae Scientiae, 18*(4), 423–447.

Schaal, N. K., Pfeifer, J., Krause, V., & Pollok, B. (2015). From amusic to musical? – Improving pitch memory in congenital amusia with transcranial alternating current stimulation. *Behavioural Brain Research, 294*, 141–148.
Schröger, E. (1997). On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology, 34*(3), 245–257.
Seesjärvi, E., Särkämö, T., Vuoksimaa, E., Tervaniemi, M., Peretz, I., & Kaprio, J. (2016). The nature and nurture of melody: A twin study of musical pitch and rhythm perception. *Behavior Genetics, 46*(4), 506–515.
Seifart, F., Meyer, J., Grawunder, S., & Dentel, L. (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society Open Science, 5*(4), 170354.
Shepard, R. N., & Metzler, D. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.
Shinozaki, J., Hiroe, N., Sato, M.-a., Nagamine, T., & Sekiyama, K. (2016). Impact of language on functional connectivity for audiovisual speech integration. *Scientific Reports, 6*, 31388.
Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage, 12*(6), 657–663.
Sorokin, A., Alku, P., & Kujala, T. (2010). Change and novelty detection in speech and non-speech sound streams. *Brain Research, 1327*, 77–90.
Sowiński, J., & Dalla Bella, S. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia, 51*, 1952–1963.
Stewart, L. (2008). Fractionating the musical mind: insights from congenital amusia. *Current Opinion in Neurobiology, 18*, 127–130.
Sun, Y., Lu, X., Ho, H. T., Johnson, B. W., Sammler, D., & Thompson, W. F. (2018). Syntactic processing in music and language: Parallel abnormalities observed in congenital amusia. *NeuroImage: Clinical, 19*, 640–651.
Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Zion Golumbic, E. (2017). Neural entrainment to the beat: The "missing-pulse" phenomenon. *Journal of Neuroscience, 37*(26), 6331–6341.
Tang, W., Wang, X.-j., Li, J.-q., Liu, C., Dong, Q., & Nan, Y. (2018). Vowel and tone recognition in quiet and in noise among Mandarin-speaking amusics. *Hearing Research*, *363*, 62–69.

Thompson, W. F. (2007). Exploring variants of amusia: Tone deafness, rhythm impairment, and intonation insensitivity. *Proceedings of the International Conference on Music Communication Science*, 2159–2063.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences, 109*, 19027–19032.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011a). Congenital Amusia (or Tone-Deafness) interferes with pitch processing in tone languages. *Frontiers in Psychology, 2*(120).

Tillmann, B., Gosselin, N., Bigand, E., & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex, 48,* 1073–1078.

Tillmann, B., Jolicoeur, P., Ishihara, M., Bertrand, O., Rossetti, Y., & Peretz, I. (2010). The amusic brain: Lost in music, but not in space. *Plos ONE, 5*, 1–6.

Tillmann, B., Lalitte, P., Albouy, P., Caclin, A. & Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia*, *85*, 10–18.

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., & Peretz, I. (2011b). Fine-grained pitch processing of music and speech in congenital amusia. *The Journal of the Acoustical Society of America*, *130,* 4089–4096.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition, 71*, 259–264

Toledo-Fernández, A., Villalobos-Gallegos, L., García-Gómez, L., & Salvador-Cruz, J. (2018). Validity of the Montreal Battery of Evaluation of Amusia: An analysis using structural equation modeling. *Studia Psychologica, 60*(1), 42–56.

Vuvan, D. T., Paquette, S., Mignault Goulet, G., Royal, I., Felezeu, M., & Peretz, I. (2018). The Montreal Protocol for Identification of Amusia. *Behavior Research Methods, 50*(2), 662–672.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research, 18*, 149–161.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiologial evidence. *Cognitive Brain Research, 25*, 180–187.

Wechsler, D. (1964). *Hamburg-Wechsler-Intelligenztest fuer Erwachsene: HAWIE* 3rd ed., Bern: Huber.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35–44.

Whiteford, K. L., & Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia, 99*, 213–224.

Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon Press.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory, 18*(6), 657–669.

Williamson, V. J., Cocchini, G., & Stewart, L. (2011). The relationship between pitch and space in congenital amusia. *Brain and Cognition, 76*(1), 70–76.

Williamson, V. J., Liu, F., Peryer, G., Grierson, M., & Stewart, L. (2012). Perception and action de-coupling in congenital amusia: Sensitivity to task demands. *Neuropsychologia, 50,* 172–180.

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology, 6*(6), 15–22.

Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2018). An electrophysiological megastudy of spoken word recognition. *Language, Cognition and Neuroscience, 33*(8), 1063–1082.

Wise, K. (2009). *Understanding "tone deafness": A multi-componential analysis of perception, cognition, singing and self-perception in adults reporting musical difficulties*. PhD dissertation, Keele University.

Ylinen, S., Huotilainen, M., & Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. *NeuroReport, 16*(16), 1857–1860.

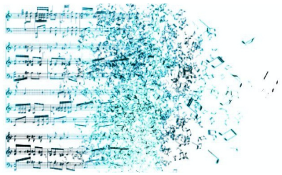
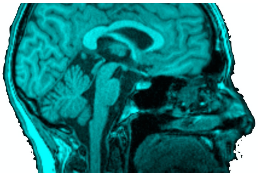
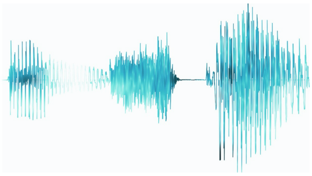
Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research, 1072*(1), 175–185.

Zhang, C., Shao, J., & Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PLoS ONE, 12*(8): e0183151.

Zhou, L., Liu, F., Jiang, J., & Jiang, C. (2019). Impaired emotional processing of chords in congenital amusia: Electrophysiological and behavioral evidence. *Brain and Cognition, 135,* 103577.

Zhou, L., Liu, F., Jiang, J., Jiang, H., & Jiang, C. (2019). Abnormal neural responses to harmonic syntactic structures in congenital amusia. *Psychophysiology, 56*, e13394.

Zhou, L., Liu, F., Jing, X., & Jiang, C. (2017). Neural differences between the processing of musical meaning conveyed by direction of pitch change and natural music in congenital amusia. *Neuropsychologia, 96*, 29–38.

8. Summary

Untwisting Amusia
What behavior, brain waves and genetic underpinnings reveal about perception in congenital amusia

In short, this thesis explores how different areas of perception are affected by congenital amusia, a disorder that negatively affects pitch and rhythm perception.

What follows is a slightly longer summary that presents the topic of this thesis and its findings in a very condensed version that is meant to be accessible to everyone.

Congenital Amusia is an innate disorder that has long been characterized as negatively affecting music perception, hence its name. This disorder is not caused by hearing loss, brain damage or an insufficient exposure to music in childhood and about 1.5% to 4% of the general population are said to be affected. Furthermore, it is assumed that amusia has a hereditary component.

Congenital amusia causes lifelong deficits in pitch and partly also rhythm perception, and the most apparent symptoms to the affected individuals themselves are various inabilities in the musical domain such as: Recognition of familiar melodies, detection of out-of-tune notes or singing or an inability to clap or sing along. Possibly due to those clear symptoms, early research has mostly focused on the influence of amusia on music perception. Hence, congenital amusia has long been characterized as a music-specific disorder.

Different aspects of musical engagement have been assessed over time and found to be impaired in amusia, such as pitch perception, pitch production, rhythm perception, beat synchronization, timbre perception and the perception of musical emotions. The underlying cause of this disorder is still unknown and it has been hypothesized to be a fine-grained pitch processing deficit, a pitch memory deficit, a statistical learning deficit or a rapid-auditory processing deficit. There is no consensus yet and it is likely a multi-causal deficit that is responsible for the different symptoms exhibited by amusics.

Regardless of the underlying deficit, pitch perception also plays an important role in language perception and recently more attention has been paid to possible pitch perception impairments in speech due to amusia. Pitch is important for the transfer of linguistic meaning. In intonation it is, for example, used to disambiguate questions from statements or for emphasis; on the word level it is used to distinguish words with similar segmental structure but different stress patterns (e.g. English *present* vs. *present*, where underscore denotes the stressed syllable) or to distinguish words with identical segments but different tones (e.g. in tone languages such as Mandarin Chinese). Other areas of speech perception have also been shown to be affected by amusia, e.g. emotional prosody in language or tone language perception. Taken together, all of these findings have evoked a change in how this disorder is seen: Previously it was described as domain-specific to music, whereas now it is viewed as a domain-general disorder affecting auditory processing.

However, there are still many unsolved questions when it comes to congenital amusia in general but in particular in regard to speech perception and in this thesis I set out to address a few of them.

In order to do so, I had to find and recruit a pool of amusic individuals to conduct tests with over the course of several years. That in itself is no easy task. Especially because we soon realized that the current test that was used to diagnose amusia, the so called *Montreal Battery of Evaluation of Amusia* (or short MBEA), had a number of shortcomings: Different numbers of subtests and different cut-off scores and different kinds of statistics were used, some research groups used web-based testing, while others didn't. This led to different prevalences and different subtypes that were found.

To resolve these issues, we conducted a large-scale study with 228 German undergraduate students who were assessed with the MBEA and a comprehensive questionnaire. The MBEA is a so-called behavioral test in which participants are auditorily presented with pairs of short piano melodies and they have to compare and judge whether the pairs are identical or not. Our experiment tested the difference between scores that were obtained in a web-based study at participants' homes and those obtained under laboratory conditions with a computerized version of the MBEA. In addition to traditional statistical procedures, we also analyzed our data with an alternative statistical procedure, called *Signal Detection Theory* (short: SDT). This statistic takes the individual's ability to discriminate something into account as well as their so-called response bias, i.e. their tendency to choose one thing rather than the other. We showed that the usage of *Signal Detection Theory* offered a more accurate way of analyzing the results of the MBEA. In addition, we showed that a diagnosis based on an average score lead to cases of misdiagnosis and we identified different subgroups of amusics in our data. We also showed that web-based testing from home resulted in many problems at least when it comes to using auditory stimuli such as the MBEA. We therefore advised against it. In short, we made several suggestions to improve the diagnosis with the MBEA that we then followed in our later studies.

After finding more reliable diagnostic standards and recruiting a pool of amusics, it was time to design and conduct several speech perception studies: After reading a lot of literature and identifying open research questions and choosing the appropriate research methods, we carefully designed several studies by recording and editing auditory stimuli, picking and then programming an experimental design, piloting it, calculating the number of necessary participants, choosing the appropriate statistics and finally running the experiment with amusic and so-called matched control participants and lastly analyzing the results (and then writing everything down, presenting it at conferences and then starting all over again).

We identified two different areas that we wanted to investigate in speech perception: On the one hand word stress perception and on the other hand vowel perception. As was already mentioned above, pitch plays an important role in speech perception, e.g. on the word level when it is used to disambiguate so called stress minimal pairs, as *present* vs. *present*, above. Regular minimal pairs in linguistics are words that are identical save for one sound, such as the word pair house-mouse. The stress minimal pairs that we used had identical segments and only differed in the stress pattern. Stress in language is manifested through different acoustic parameters. The two most important ones that we investigated in our word stress studies were pitch and duration. However, speech also makes extensive use of other information in the speech signal such as spectral frequencies. The latter are especially relevant in the perception of vowels. For our vowel study, we used such spectral contrasts as for example between the English vowels in *heed* versus *hid* and also durational contrasts.

We conducted a total of four studies about speech perception: Two so-called behavioral studies preceding two electrophysiological studies. In the behavioral studies, our participants were invited to our laboratory, seated in a sound-insulated booth and presented with pairs (or triplets) of auditory stimuli on a computer and they had to decide whether the sounds were identical or which two out of three sounds were most similar to one another. Analyzing these results helped us to find out what amusics could consciously perceive.

In the next step, we were then interested in what amusics perceived unconsciously. In this case 'unconsciously' does not refer to the participants being unconscious but rather to the fact that they did not have to pay conscious attention to any task, while we recorded their brainwaves (the so-called electroencephalogram (EEGs)) while they were listening to sounds. After attaching around 70 electrodes to their head and face, participants were again seated in our sound-attenuated booth. This time they watched a muted nature documentary and heard our sound stimuli to which they did not have to pay attention. Their only task was not to move around too much and not to fall asleep, which is both difficult enough under those conditions. In the meantime, we recorded the event-related potentials (short: ERPs), i.e. the measurable brain responses to outside stimuli, which is a noninvasive way of measuring brain activity. We were interested in one of many components in particular, the mismatch negativity (MMN), which is a very early component that is evoked by unconscious change detection, i.e. it is the brains reaction to an odd stimulus in a sequence of stimuli, which does not need to be consciously registered.

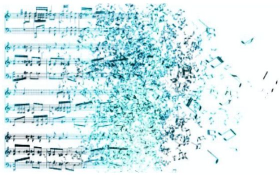
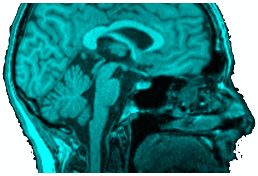
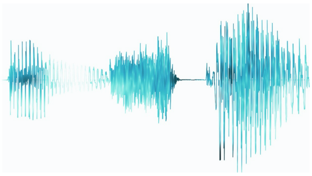
Taken together, these studies revealed that amusics consciously struggled in discriminating vowel pairs based on spectral as well as durational cues, while durational cues were overall harder for both groups.

In the word stress study, amusics revealed no difficulties concerning pitch processing and a better usage of durational cues. The MMN responses for both studies were the same however: Amusics had an MMN, which is a finding in and off itself, however it was significantly reduced in comparison to controls. These contrasting results call for a more in-depth investigation of further MMN components and behaviorally of possible compensation strategies that amusics might have developed.

Lastly, while building our testing pool, we found a dizygotic twin pair, of whom one twin is amusic and the other one is not. This posed an interesting case to investigate, as the twins grew up together, went to school together and their exposure to music throughout their lives was as comparable as it can be for two individuals. Yet one is amusic while the other one is not. This twin case study proved that congenital amusia is not due to insufficient exposure to music in childhood.

We conducted a large battery of tests assessing the performance of the twins in music, pitch perception and memory, language perception and spatial processing. In addition to the expected differences in melodic and language perception abilities, we found that both twins exhibit a comparably low pitch memory span and low beat perception abilities. This raised questions of nature versus nurture in congenital amusia, i.e. whether their shared genes or their shared environment and low musical education is responsible for the shared deviant performance. This in turn gives rise to the question of hereditariness of congenital amusia and calls for a genetic analysis of affected individuals.

All in all, the goal of this thesis was to untangle certain aspects of auditory perception in congenital amusia by investigating its cognitive, neural and genetic underpinnings. While some questions have received answers, many new questions arose.

9. Samenvatting

Amusie ontrafelen
Wat gedrag, hersengolven en erfelijkheid onthullen over perceptie bij aangeboren amusie

Om kort te gaan: Dit proefschrift gaat over de invloed van congenitale amusie, een aandoening die toonhoogte en ritmewaarneming negatief beïnvloedt op verschillende gebieden van auditieve waarneming.

Nu volgt een meer gedetailleerde samenvatting van dit proefschrift, waarin de verschillende experimenten kort worden geschetst voorafgegaan door een beschrijving van amusie.

Congenitale amusie is een aangeboren waarnemingsstoornis die ervoor zorgt dat getroffenen problemen hebben met muziek- en spraakperceptie, omdat toonhoogte en ritmeperceptie kunnen worden verstoord. Deze neurologische aandoening is aangeboren. Ze kan tot nu toe niet worden behandeld, en het exacte onderliggende mechanisme is nog onbekend. Amusie wordt niet veroorzaakt door gehoorbeschadiging, hersenbeschadiging of onvoldoende blootstelling aan muziek in de kindertijd, en ongeveer 1,5% - 4% van de algemene bevolking lijden eraan.

Mensen met amusie hebben hun hele leven moeite met de waarneming van toonhoogte en ritme. De symptomen van amusici zijn het duidelijkst op het gebied van muziek. Deze omvatten het onvermogen om een bekende melodie zonder tekst te herkennen, het onvermogen om te herkennen wanneer iemand vals zingt, het onvermogen om noten van verschillende toonhoogte of timbre te onderscheiden, het onvermogen om geluiden of melodieën correct te (her)produceren en het onvermogen om ritmes te produceren of te onderscheiden.

Vanwege deze zeer duidelijke symptomen concentreerde het vroege amusie-onderzoek zich voornamelijk op de muzikale aspecten, daarom wordt amusie vaak een muziekspecifieke stoornis genoemd. Omdat toonhoogtewaarneming echter ook een belangrijke rol speelt in de perceptie van spraak, is het onderzoek naar taalkundige aspecten de laatste jaren belangrijker geworden. Als we spreken, gebruiken we toonhoogteverschillen, bijvoorbeeld in intonatie, om vragen van uitspraken te onderscheiden of om iets te benadrukken. Inmiddels wordt amusie niet meer gezien als een muziekspecifieke stoornis, maar als een gebiedsoverschrijdende stoornis die de auditieve waarneming aantast.

Er zijn echter veel onbeantwoorde vragen over aangeboren amusie. In dit proefschrift heb ik er een aantal proberen te beantwoorden.
Om dit te kunnen doen, moest ik eerst een groep van amusici vinden bij wie ik mijn verschillende studies kon uitvoeren. Dat was geen gemakkelijke taak, omdat amusici meestal niet weten dat ze amusie hebben. We moesten daarom eerst mensen uitnodigen voor een diagnostische test in ons laboratorium via flyers met vragen over hun relatie met muziek. Bovendien merkten we al snel dat de enige test voor diagnose, de zogenaamde Montreal Battery of Evaluation of Amusia (of kortweg MBEA), tekortkomingen vertoont in toepassing en evaluatie.

Om deze tekortkomingen verder te onderzoeken en suggesties te kunnen doen voor verbeteringen, hebben we een grootschalig onderzoek gedaan onder 228 studenten. De MBEA is een zogenaamde gedragstest waarbij de deelnemers twee korte pianomelodieën horen en met elkaar vergelijken en beslissen of de melodieën hetzelfde of verschillend waren. In ons onderzoek hebben we gekeken of er een verschil is wanneer de test in het laboratorium wordt uitgevoerd of alleen online thuis, of er verkeerde diagnoses zijn en of de juiste statistieken worden gebruikt. We konden aantonen dat online tests niet betrouwbaar waren, dat er mogelijk verkeerde diagnoses waren en dat andere statistieken beter geschikt zijn voor evaluatie. Op grond daarvan deden we verschillende aanbevelingen om de toepassing en evaluatie van de MBEA te verbeteren, die we vervolgens zelf hebben gevolgd. Dit alles is terug te vinden in het 2e hoofdstuk van dit proefschrift.
Nadat we betrouwbaardere diagnostische normen voor onszelf hadden vastgesteld en een pool van amusici hadden gerekruteerd, was het tijd om onze waarnemingsstudies uit te voeren.

We wilden twee verschillende gebieden in de spraakperceptie van amusici onderzoeken. Enerzijds de perceptie van klinkers en anderzijds de perceptie van klemtoon. Zoals hierboven vermeld spelen toonhoogteverschillen een belangrijke rol. Op woordniveau gebruiken we toonhoogteverschillen en onder andere lengte om woorden met dezelfde klanken maar verschillende klemtoonpatronen van elkaar te onderscheiden, bijvoorbeeld z.B. *canon* vs. *kanon* (de onderstreping toont de klemtoon). Er zijn andere akoestische parameters waarmee we *accentuering* markeren, maar in onze studie in hoofdstuk 4 hebben we ons gericht op verschillen in toonhoogte en lengte en welke invloed deze hebben op de perceptie van amusici.

Het spraaksignaal is echter erg complex, daarom hebben we ons in een andere studie gericht op de rol van de waarneming van spectrale frequenties in de perceptie van klinkers in hoofdstuk 3. Deze spectrale frequenties onderscheiden de verschillende klinkers van elkaar, bijvoorbeeld in de woorden *mest en meest*.

Op deze twee gebieden hebben we in totaal vier onderzoeken uitgevoerd. Twee studies waarbij de deelnemers plaats moesten nemen in de geluiddichte cabine in ons laboratorium en moesten luisteren naar paren van stimuli, woorden of individuele klinkers, en ze moesten beslissen of deze paren hetzelfde of verschillend waren. Dit moet ons helpen iets te weten te komen over de bewuste waarneming van amusici.

Om ook iets over onbewuste waarneming te weten te komen, hebben we in de volgende stap twee onderzoeken gedaan waarin we de hersengolven van de deelnemers als reactie op onze stimuli met elektro-encefalografie (EEG) hebben gemeten.

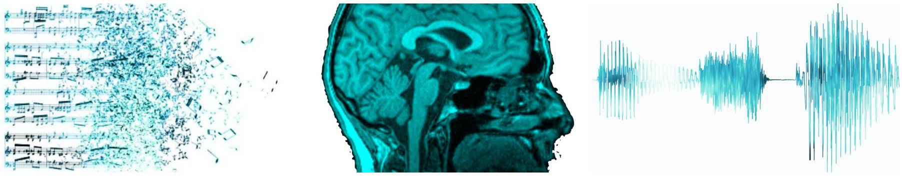
Alles bij elkaar genomen hebben deze onderzoeken aangetoond dat amusici moeite hebben met het bewust waarnemen van de verschillen tussen klinkerparen op basis van hun lengte en op basis van hun spectrale frequenties. In de klemtoonstudie hadden amusici er geen moeite mee om woordparen bewust te onderscheiden op de basis van de verschillen in toonhoogte en lengte. De EEG-onderzoeken hadden echter vergelijkbare resultaten: amusici reageerden onbewust op de stimuli. Dit is op zich een nieuwe bevinding. Hun hersengolven als reactie op de stimuli waren echter aanzienlijk verminderd in vergelijking met de controlepersonen.

Deze contrasterende resultaten laten zien dat verder onderzoek nodig is om erachter te komen welke mogelijke compensatiestrategieën amusici hebben ontwikkeld.

In het laatste artikel, in hoofdstuk 5, hebben we een tweeling onderzocht, waarvan de ene amuzikaal is en de andere niet. Dit was een zeer interessant geval, aangezien ze allebei samen opgroeiden, dezelfde schoolopleiding genoten en in hun jeugd in dezelfde mate werden blootgesteld aan muziek. Hun leven was dus zo vergelijkbaar als het van twee mensen zou kunnen zijn. Toch is de ene amuzikaal en de andere niet. We voerden een grote reeks tests uit op de tweelingen en testten hun prestaties op het gebied van muziekperceptie, waarneming van toonhoogteverschillen, spraakperceptie en geheugen.

We konden aantonen dat de tweeling de verwachte verschillen in muziek- en spraakperceptie had. Beiden vertoonden echter een beperkt geheugen voor toonhoogteverschillen en een beperkt geheugen vor ritmische verschillen. Hieruit blijkt dat er dringend behoefte is aan verder onderzoek naar de erfelijkheid van amusie.

In het algemeen was het doel van dit proefschrift om enkele onverklaarde aspecten van auditieve waarneming bij amusie te ontrafelen. Cognitieve en neurale processen en erfelijkheid werden onderzocht. Hoewel sommige vragen zijn beantwoord, leiden deze antwoorden tot nieuwe vragen, die verder onderzocht moeten worden.



10. Zusammenfassung

Amusie entwirren
Was Verhalten, Hirnströme und Genetik über Wahrnehmung in kongenitaler Amusie verraten

Diese Dissertation befasst sich mit dem Einfluss kongenitaler Amusie auf das Hören und die Verarbeitung des Gehörten im Gehirn. Bei *kongenitaler Amusie* handelt es sich um eine Störung, die dazu führt, dass Menschen Tonhöhen und Rhythmen nicht so gut wahrnehmen können, wie andere Menschen ohne diese Störung.

Es folgt nun eine etwas ausführlichere Zusammenfassung dieser Dissertation, mit folgendem Aufbau: Als erstes wird Amusie genauer beschrieben. Dann wird dargestellt, wie wissenschaftliche Studien generell entstehen und zum Schluss werden die Ergebnisse der verschiedenen Experimente vorgestellt, die für diese Dissertation durchgeführt worden sind.

Kongenitale Amusie ist eine angeborene Wahrnehmungsstörung, die dafür sorgt, dass Betroffene Probleme bei Musik- und Sprachwahrnehmung haben, da Tonhöhen- und Rhythmuswahrnehmung gestört sein können. Die Störung ist von Geburt an vorhanden. Es wird angenommen, dass sie vererbt wird; sie ist neurologisch im Ursprung und sie ist bisher nicht behandelbar. Außerdem ist das genaue zugrundeliegende Defizit bisher unbekannt. Gesichert ist aber, woher Amusie nicht kommt: Amusie wird weder durch eine Hörschwäche, noch durch Hirnschädigung und auch nicht durch ungenügenden Kontakt zu Musik in der Kindheit ausgelöst. Ca. 1,5% - 4% der allgemeinen Bevölkerung sind davon betroffen – sehr oft, ohne es zu wissen.

Betroffene haben ihr gesamtes Leben lang Schwierigkeiten bei der Tonhöhen- und Rhythmuswahrnehmung. Am meisten fallen den Betroffenen Symptome im Umgang mit Musik auf: Sie können z.B. selbst sehr bekannte Melodien nicht ohne dazugehörigen Liedtext erkennen; auch können sie „schiefes Singen“ nicht erkennen und ebenso wenig Noten unterschiedlicher Tonhöhe oder Klangfarbe unterscheiden. Sie sind nicht in der Lage, Töne oder Melodien korrekt wiederzugeben sowie Rhythmen zu produzieren oder zu unterscheiden.

Durch diese recht klaren Symptome hatte sich die frühe Amusieforschung hauptsächlich auf musikalische Aspekte konzentriert,

weswegen Amusie lange als musikspezifische Störung bezeichnet worden ist. Da aber Tonhöhenwahrnehmung auch eine wichtige Rolle in der *Sprachwahrnehmung* spielt, wurde in den letzten Jahren auch verstärkt an linguistischen Aspekten geforscht. Wenn wir sprechen, benutzen wir Tonhöhenunterschiede z.B. um Fragen von Aussagen zu unterscheiden oder um etwas besonders zu betonen. Mittlerweile wird Amusie daher nicht mehr als musik-spezifische Störung gesehen, sondern als eine übergreifende Störung, die die auditive Wahrnehmung betrifft.
Es gibt allerdings weiterhin viele unbeantwortete Fragen zu kongenitaler Amusie. In dieser Dissertation habe ich versucht, ein paar davon zu beantworten.

Um das tun zu können, musste ich als erstes eine Gruppe von AmusikerInnen rekrutieren mit denen ich dann meine verschiedenen Studien durchführen konnte. Das allein war keine einfache Aufgabe, da AmusikerInnen meistens nicht wissen, dass sie Amusie haben. Wir mussten sie deshalb erst durch Flyer mit Fragen zu ihrem Verhältnis zu Musik zu einem Diagnosetest in unser Labor einladen. Außerdem bemerkten wir bald, dass der einzige Test zur Diagnose, die sogenannte *Montreal Battery of Evaluation of Amusia* (or kurz MBEA) Mängel in Anwendung und Auswertung aufweist. Um diese Mängel weiter zu untersuchen und Vorschläge zur Verbesserung der Diagnose machen zu können haben wir eine groß angelegte Studie mit 228 Studierenden durchgeführt. Die MBEA ist ein sogenannter Verhaltenstest, bei dem die TeilnehmerInnen zwei kurze Klaviermelodien hören. Sie sollen diese dann miteinander vergleichen und entscheiden, ob die Melodien gleich oder unterschiedlich waren.

In unserer Studie haben wir untersucht, ob es einen Unterschied macht, ob der Test im Labor oder alleine zuhause online durchgeführt wird, ob es zu Fehldiagnosen kommt und ob die passenden Statistiken angewendet werden.
Wir konnten zeigen, dass Tests über das Internet nicht verlässlich waren, dass es zu möglichen Fehldiagnosen kam und dass die zur Auswertung benutzten Statistiken nicht so gut geeignet sind und es bessere Alternativen gibt. Wir haben so mehrere Verbesserungsvorschläge zur Diagnose von Amusie machen können, die wir dann auch selbst befolgt haben. Das alles ist im 2. Kapitel dieser Doktorarbeit nachzulesen.

Nachdem wir auf diesem Weg eine verlässlichere Diagnose von Amusie ermöglicht und genügen AmusikerInnen gefunden hatten, war es an der Zeit, unsere eigentlichen Wahrnehmungsstudien durchzuführen. Im Folgenden beschreibe ich, wie der generelle Prozess der Planung, Durchführung und Auswertung einer wissenschaftlichen Studie funktioniert und welche Schritte für jede einzelne Studie durchlaufen werden müssen und wurden:

Als erstes habe ich sehr viele wissenschaftliche Bücher und Fachzeitschriften gelesen und offene Fragestellungen identifiziert. Dann haben wir angemessene Forschungsmethoden ausgewählt und unsere Studien geplant und entworfen. Dafür haben wir Audioaufnahmen gemacht und diese mit dem Computerprogramm Praat bearbeitet, so dass daraus unsere sogenannten Stimuli, also die Testwörter des Experiments, entstanden sind. Dann haben wir unser experimentelles Paradigma, also die genaue Aufgabenstellung, ausgewählt und es ebenfalls in Praat programmiert. Außerdem muss ein sogenanntes Ethikvotum eingeholt werden, welches uns erlaubt hat, die Studien durchzuführen, weil sie ethisch in Ordnung waren. Als nächstes wurden dann sogenannte Pilotstudien durchgeführt. Dabei wird vor dem eigentlichen Studienbeginn getestet, ob ein Experiment funktioniert, ob die Programmierung funktioniert und die Aufgabenstellung verständlich ist. Vor Beginn jeder Studie wurde ausgerechnet, wie viele Personen daran mindestens teilnehmen müssen und es wurde festgelegt, welche statistischen Tests geeignet und angemessen zur Auswertung sind. Dann wurden die Studien im Labor durchgeführt, in dem eine ausreichende Anzahl an AmusikerInnen und sogenannten Kontrollpersonen teilnahm. Die Kontrollpersonen stimmten in vielen Eigenschaften mit den AmusikerInnen überein – außer eben in ihrer Amusie. Das war wichtig, um Vergleiche zwischen AmusikerInnen und Personen ohne Amusie anstellen zu können. Nachdem die Studie durchgeführt worden war und die Daten alle an mehreren Orten gespeichert worden sind, konnten die Ergebnisse statistisch ausgewertet werden. Die Ergebnisse wurden auf internationalen Fachtagungen präsentiert und dann wurde alles aufgeschrieben, bei Fachzeitschriften eingereicht, von internationalen Experten begutachtet und dann als Artikel veröffentlicht. Diese Fachartikel sind – bis aus Einleitung und Fazit – die einzelnen Kapitel dieser Dissertation.

Wir haben zwei verschiedene Bereiche in der Sprachwahrnehmung von AmusikerInnen identifiziert, die wir untersuchen wollten: Zum einen die

Vokalwahrnehmung und zum anderen die Wortbetonungswahrnehmung. Wie oben bereits erwähnt, spielen Tonhöhenunterschiede eine wichtige Rolle. Auf der Wortebene benutzen wir Tonhöhenunterschiede und unter anderem Länge, um Wörter mit gleichen Segmenten, aber unterschiedlichen Betonungsmustern voneinander zu unterscheiden, z.B. *umreiten* vs. *umreiten* (die Unterstreichung zeigt die Betonung). Es gibt noch weitere akustische Aspekte, mit denen wir Betonung markieren, aber wir haben uns in unserer Studie in Kapitel 4 auf Tonhöhenunterschiede und Länge konzentriert und darauf, welchen Einfluss sie auf die Wahrnehmung von Amusikern haben. Das Sprachsignal ist jedoch sehr komplex, daher haben wir uns in unserer weiteren Studie in Kapitel 3 zu Vokalwahrnehmung auf die Wahrnehmung von spektralen Frequenzen konzentriert. Diese Frequenzen und wiederum Länge unterscheiden die verschiedenen Vokale, z.B. in den Wörtern *B**e**tten*, *b**e**ten, und b**ä**ten,* voneinander.

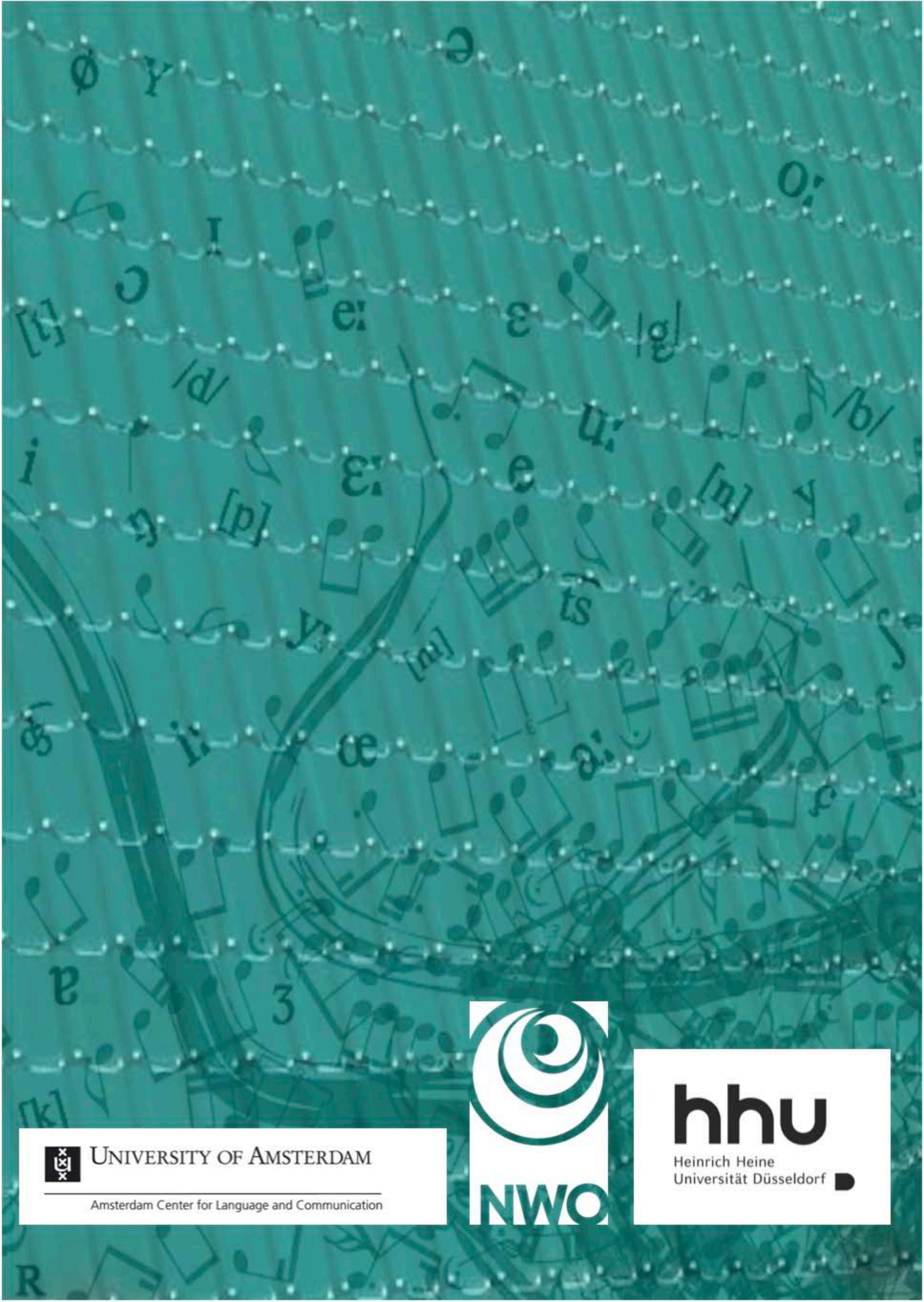
Zu diesen beiden Bereichen haben wir insgesamt vier Studien durchgeführt. Zwei Verhaltensstudien in denen die TeilnehmerInnen im Labor in unserer schallisolierten Kabine Platz nehmen mussten und Paare von Stimuli - Wörter oder einzelne Vokale - anhören mussten und entscheiden mussten, ob diese gleich oder unterschiedlich waren. Dies sollte uns helfen, etwas über die bewusste Wahrnehmung von AmusikerInnen herauszufinden. Um ebenfalls etwas über die unbewusste Wahrnehmung herauszufinden haben wir im nächsten Schritt zwei Studien durchgeführt, in denen wir mittels Elektroenzephalografie (EEG) die Gehirnströme der TeilnehmerInnen in Reaktion auf unsere Stimuli aufgezeichnet haben.

Zusammengefasst haben diese Studien gezeigt, dass AmusikerInnen in ihrer bewussten Wahrnehmung Probleme haben, Vokalpaare basierend auf ihrer Dauer als auch basierend auf ihren spektralen Qualitäten zu unterscheiden. In der Wortbetonungsstudie hatten AmusikerInnen keine Probleme in der bewussten Unterscheidung von Wortpaaren aufgrund der Tonhöhen- und Längenunterschiede. Die EEG-Studien hatten jedoch vergleichbare Ergebnisse: AmusikerInnen hatten eine unbewusste Reaktion auf die Stimuli. Dies allein ist schon eine neue Erkenntnis. Jedoch waren ihre Gehirnwellen als Reaktion auf die Stimuli deutlich vermindert im Vergleich zu den Kontrollpersonen. Diese kontrastierenden Ergebnisse zeigen, dass weitere Studien angebracht sind, um herauszufinden, welche möglichen Kompensationsstrategien AmusikerInnen entwickelt haben.

Im letzten Aufsatz, in Kapitel 5, haben wir ein Zwillingspaar untersucht. Eine Zwillingsschwester ist amusisch, die andere nicht. Dies war wissenschaftlich ein sehr interessanter Fall, da beide zusammen aufgewachsen waren, die gleiche Schulausbildung hatten und in gleichem Umfang in ihrer Kindheit mit Musik in Berührung gekommen sind. Ihre Leben waren also so vergleichbar, wie es bei zwei Individuen nur sein kann. Trotzdem ist eine Zwillingsschwester amusisch und die andere nicht. Wir haben eine große Bandbreite an Tests mit den beiden Zwillingen durchgeführt und ihre Leistung in Musikwahrnehmung, Tonhöhenunterscheidung, Sprachwahrnehmung und Gedächtnis getestet.

Wir konnten zeigen, dass die Zwillinge die erwarteten Unterschiede in Musik- und Sprachwahrnehmung aufwiesen. Allerdings zeigten beide Zwillinge eine eingeschränkte Gedächtnisleistung für Tonhöhenunterschiede und eine eingeschränkte Taktunterscheidung. Diese Studie zeigt, dass dringend eine weitere Beschäftigung mit der Vererblichkeit von Amusie notwendig ist, weil hier noch viele Fragen offen sind

Generell war es das Ziel dieser Arbeit einige ungeklärte Aspekte der auditiven Wahrnehmung in Amusie zu entwirren. Es wurden dabei kognitive und neuronale Vorgänge sowie Vererblichkeit erforscht. Während einige Fragen durch die Arbeit beantwortet worden sind, sind weitere neue dabei entstanden, die erforscht werden wolle



UNIVERSITY OF AMSTERDAM

Amsterdam Center for Language and Communication

NWO

hhu

Heinrich Heine
Universität Düsseldorf