Word stress perception by congenital amusics

Pfeifer, J.; Hamann, S.

DOI
10.1515/9783110770186-005

Publication date
2022

Document Version
Final published version

Published in
How Language Speaks to Music

Jasmin Pfeifer, Silke Hamann

Word stress perception by congenital amusics

Abstract: 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 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 fundamental frequency (pitch), duration, intensity and spectral slope. Behavioral results surprisingly revealed no pitch processing difficulties for word stress in the amusic group in comparison to controls, and amusics also showed a better usage of durational cues. The electrophysiological results revealed that amusics consistently have an MMN, though it is smaller than that of controls. The present results warrant further investigation of the use of linguistic cues by congenital amusics.

Keywords: Congenital Amusia, Word Stress, Pitch, Duration, MMN

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 et al., 2004; Stewart, 2008). 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

Acknowledgements: This work was supported by the *Netherlands Organisation for Scientific Research* (NWO; grant number 322-75-004) to the first author.

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

Jasmin Pfeifer, Institute for English Language and Linguistics, Heinrich-Heine-University, Germany, e-mail: pfeifer@phil.hhu.de
Silke Hamann, Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands, e-mail: silke.hamann@uva.nl

https://doi.org/10.1515/9783110770186-005

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 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 et al., 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 et al., 2011), rhythm perception (Foxton et al., 2006), beat synchronization (Sowiński & Dalla Bella, 2013), timbre perception (Marin et al., 2012a), consonance rating (Ayotte et al., 2002), and musical emotion perception (Marin et al., 2012b).

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 et al., 2010; Hyde & Peretz, 2004), a pitch memory deficit (Gosselin et al., 2009; Tillmann et al., 2009; Williamson & Stewart, 2010; Tillmann et al., 2016), a statistical learning deficit (Peretz et al., 2012) or a rapid-auditory processing deficit (Williamson et al., 2010; Albouy et al., 2016) and partly also a rhythm/beat perception deficit (Phillips-Silver et al., 2013; Launay et al., 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 et al., 2006); more grey matter in the right inferior frontal gyrus and right superior temporal gyrus (STG) and other brain areas (Hyde et al., 2007); less grey matter in the left IFG (Broca's area) and left STG (Wernicke's area; Mandell et al., 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 et al., 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 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) for the evaluation of MBEA results, 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 et al., 2015a; Liu et al., 2015b; Liu et al., 2012; Tillmann et al., 2011) emotional prosody in language (Lolli et al., 2015; Thompson et al., 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 et al., 2017).

The speech perception studies with most relevance to the present study are on 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 statement-question pairs that were 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 contained only vowels and voiced consonants (resulting in a continuous pitch), or they also contained voiceless consonants (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. The study by Huang et al. (2015) considered durational differences and linguistic tone at the same time, and was conducted with native speakers of a tonal language. They 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, by Jasmin et al. (2019), investigated the cue weighting of durational versus pitch cues of amusics. They investigated 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 the fundamental frequency of the following vowel as the secondary cue of pitch. 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 differences 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 the durational VOT cues than on the vowel-inherent 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 et al., 2006; Leong et al., 2011; Goswami et al., 2013), children at risk of SLI (e.g. Gallon et al., 2007; Haake et al., 2013; Fikkert & Penner, 1998) and people with Down syndrome (e.g. Pettinato & Verhoeven, 2008). 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 et al., 1978), for reviews, see Näätänen (2001); Näätänen et al. (2007); Picton et al. (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 et al., 2013; Kirmse et al., 2008; Näätänen, 2001; Partanen et al., 2011; Ylinen et al., 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 et al., 2000; Sorokin et al., 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 et al., 1998) specifically the dorsolateral prefrontal cortices were implicated. These are also the regions that are affected by amusia, as shown by Albouy et al. (2013a) and Hyde et al. (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 Goulet 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 these studies, 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 length of the vowel or syllable, with stressed vowels and syllables having longer duration than their unstressed counterparts, as shown by Dogil (1995), Jessen et al. (1995) and Haake et al. (2013), amongst others. Secondary cues are pitch, loudness 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 quite variable within and across speakers.

With respect to where the word stress is positioned and therefore expected in German, the penultimate syllable is considered the default stress position (Eisenberg, 1991; Wiese, 2000: 180–182).

1.5 Cue weighting and the relation between speech and music

The perception of speech and that of music involves the integration of several acoustic characteristics as perceptual cues – as indicated above for the case of word stress in German: mainly duration (perceived as length), fundamental frequency (perceived as pitch), intensity (perceived as loudness) and spectral slope (perceived as a combination of vowel quality and loudness). These cues have to be weighted and integrated in order to be mapped onto a categorical representation. For speech perception, spectral and temporal variations are mapped onto linguistic units, whereas in music, tonal and rhythmical variation are mapped onto musical units (e.g. Patel, 2003). This mapping and categorizing is no easy feat as it requires a very precise and rapid detection and integration of a number of acoustic cues. Lisker (1986), for example, lists 16 acoustic properties that can be used as cues to differentiate between voiced and voiceless sounds alone, i.e., the difference between e.g. /b/ and /p/. While this means that a great level of redundant information is present that needs to be integrated, those redundant cues can also ensure that a selective perceptual deficit does not make speech perception impossible, thereby making it more robust (Patel et al., 2008). Congenital amusia, with its very specific perceptual deficits and yet seemingly unimpaired overall language perception and production, offers a unique window into possible different cue weighting and integration strategies that the redundancy in the acoustic signal affords. And as detailed in the previous subsections, stress assignment in German seems to be an ideal testing ground for examining general auditory deficits and potential speech perception deficits of amusics, as it involves not only individual perceptual cues (such as pitch, length, loudness, and vowel quality) but also higher-order processes of cue weighting and integration.

2 Materials and methods

In the following subsections, we describe the stimuli (section 2.1), participants (section 2.2) and the procedure (section 2.3) of our behavioral and MMN experiment.

2.1 Stimuli

In order to test the role of duration, which was found to be the most important cue for stress perception in German (recall section 1.4), and the role of pitch, which should pose the greatest difficulty to amusics due to their pitch perception impairment (recall sections 1.1 and 1.2) on stress perception in German, we need to manipulate real speech. Natural speech always includes all potential cues, and in order to systematically test the weight of one or two of them, all other cues need to be changed to intermediate values between a stressed and an unstressed syllable, which can only be achieved by manipulation.

For the creation of the stimuli for both our experiments, we recorded a native speaker of German producing several repetitions of the stress minimal pair *umstellen* [ˈʊmʃtɛln̩] "to reposition" vs. *umstellen* "to surround" [ʊmˈʃtɛln̩] in isolation (where underscore denotes the syllable with main stress). We then picked the productions that were pronounced the clearest. Though we had many more stress minimal pairs with two distinct meanings (e.g. *untergraben*, *durchreisen*) included in our recordings and the pilot, we had to restrict the actual experiment to one pair as it would have been much too long otherwise.

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 the sequence [tɛln̩]. For each part of each word, we measured the acoustic parameters duration, fundamental frequency (which we will call by its perceptual correlate pitch in the following), spectral slope and intensity, as given in Table 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.

Table 1: Acoustic parameters of the first and second part of the two words.

	umstellen		***umstellen***	
	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

As a 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.

Since **duration** was found to be the main cue to stress perception in German, 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 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.

Table 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.

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

Of major importance for amusics is their possible impaired perception of **pitch** in speech. We therefore manipulated the pitch of our stimuli by using the natural contours of the two words and creating 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. This yielded in total three pitch contours, see Table 2, rows three to five.

With respect to the secondary stress cue **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 six and seven.

Intensity is mentioned as another relevant secondary cue of German stress perception in the literature. We therefore manipulated the intensity of the two parts of our stimuli, based again on the measures of the original recordings. In addition to these two, we also created an in-between intensity contour; see the last three rows in Table 2.

The above-described manipulation in duration, pitch, slope and intensity 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 they were consistently categorized as having stress either on the first or on the second word part, which was the case. A task similar to the main study (described below in 2.3) was used for this pilot.

For the EEG experiment, only four of the stimuli from the behavioral task were used: due to the experimental constraints of an oddball task. Two of these were the two re-synthesized endpoint stimuli with a contrast in all cues. The choice of the other two was based on the expected pitch-deficit in speech of our amusic group. For the EEG study, we therefore chose two further stimuli that contained a contradiction between pitch and all other cues: 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.

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 for each study separately in the subsections 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 study. Their characteristics can be found in Table 3, and their MBEA scores for the four relevant subtests in Table 4. The discriminatory ability of the two groups is significantly different, as shown by the *t*-test values in Table 4.

Table 3: Subject characteristics of behavioral study: 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.

Group		Age	Years of education	Years of music education	Gender
Amusics	Mean	27.43	18.86	3.00	2 male
Controls	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: Montreal Battery of Evaluation of Amusia scores for behavioral study: 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.

			Scale	Contour	Interval	Rhythm	Meter	Memory
Proportion 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*	−5.497	−9.946	−10.077	−7.104	−5.746	−4.378
		p	0.000	0.000	0.000	0.000	0.000	0.001

2.2.2 MMN experiment

10 controls and 10 amusics were included in the MMN study. Their characteristics can be found in Table 5, and their MBEA scores for the four relevant subtests in Table 6. Again, the discriminatory ability of the two groups is significantly different, as shown by the *t*-test values in Table 6. The seven amusics who participated in the behavioral study also participated in the MMN study with three additional amusics. As both studies were conducted more than 12 months apart, we do not

Table 5: Subject characteristics of MMN study: 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.

Group		Age	Years of education	Years of music education	Gender
Amusics	Mean	36.70	19.10	2.40	3 male
Controls	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 6: Montreal Battery of Evaluation of Amusia scores for MMN study: 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.

			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*	−7.253	−10.542	−10.453	−5.094	−4.924	−4.177
		p	0.000	0.000	0.000	0.000	0.000	0.001

think that there were any carryover effects for the seven participants who participated in both studies.

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 (see Summerfield, 1981). Participants were presented with a forced-choice identification task in three blocks (of 81 stimuli each) 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 indicate the meaning of the word. The two pictures that were used as answer categories were introduced to the participants with two sentences before the start of the experiment. They were hand-drawn scenes depicting typical situations for both meanings and they were piloted before the experiment. The order of these two pictures was counterbalanced between blocks and participants. Each stimulus was repeated three times and occurred once per block, and the presentation was pseudo-randomized. 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 twelve 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 et al., 2016; Shinozaki et al., 2016). Participants listened passively to the stimuli 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 contained 1800: (tokens) auditory stimuli. Each block was repeated once, resulting in four blocks in total, which were counterbalanced across participants.

The standard 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 word part or on the second word part. 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 (types) were used.

As is usual in a multi-deviant oddball paradigm, 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). The combination of 2 standards and 6 deviants yielded a total of 8 event-related potentials (ERPs) per participant. 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 preprocessing

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, 2019). 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 answers "stress on the first word part" was taken as the dependent variable. Following established modeling practices (Baayen et al., 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 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, $z = 0.98$, $p = 0.33$).

Table 7: Final generalized linear mixed model fit by maximum likelihood. Asterisks indicate significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

	Estimate	Std. Error	*z*-value	Pr(>|z|)
(Intercept)	−0.76	0.16	−4.73	2.22e-06 ***
Group	−0.25	0.18	−1.35	0.17774
Duration	1.15	0.36	3.24	0.00118 **
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 x Duration by Pitch	−0.31	0.74	−0.42	0.67724

A visualization of the observed data can be found in Figures 1 and 2. Figure 1 depicts a clear effect of pitch: Both amusics and controls identified an early pitch rise (the first two bars in each plot) as stress on the first word part (the blue bar is higher than the green one), and a late pitch rise (the last two bars in each plot) as stress on the second part (the green bar is higher than the blue one). The ambiguous stimuli with two pitch peaks (the two bars in the middle of each plot) were mainly identified as having stress on the second part.

Figure 2 depicts the effect of duration of the first vowel. All participants showed a preference to identify all stimuli as being stressed on the second word part. However, while the responses of the control group (upper panel) are

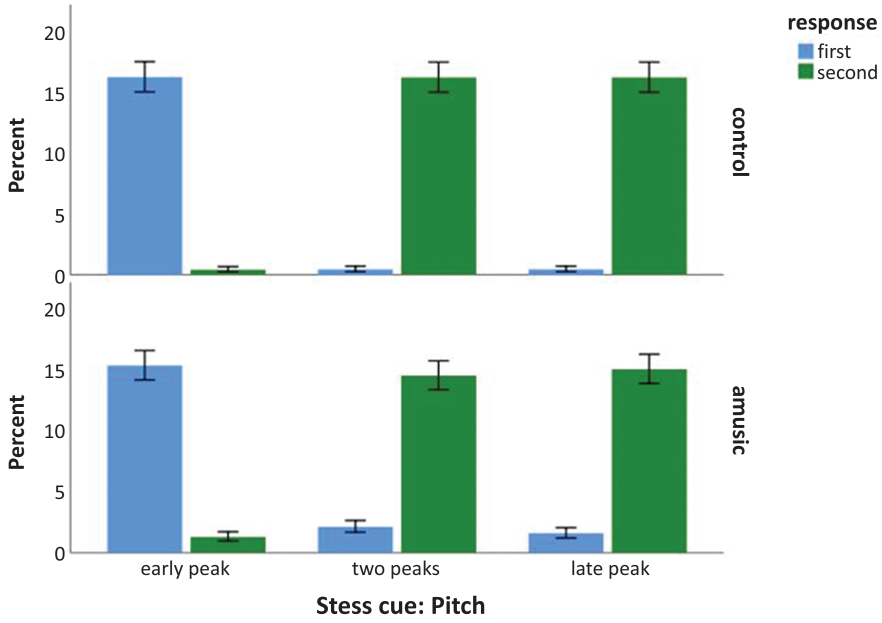

Figure 1: Responses to pitch cue split by group. Response first (blue) or second word part (green) in percent (grand total) to stimuli where pitch has an early peak (left), two peaks (middle), or a late peak (right). Top: control group, bottom: amusic group. Error bars show 95% CI.

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 part" identifications for stimuli with a long than with a mid or short first vowel. The amusics thus use durational cues more reliably than controls to identify stress.

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 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 (i.e., the waveform of the

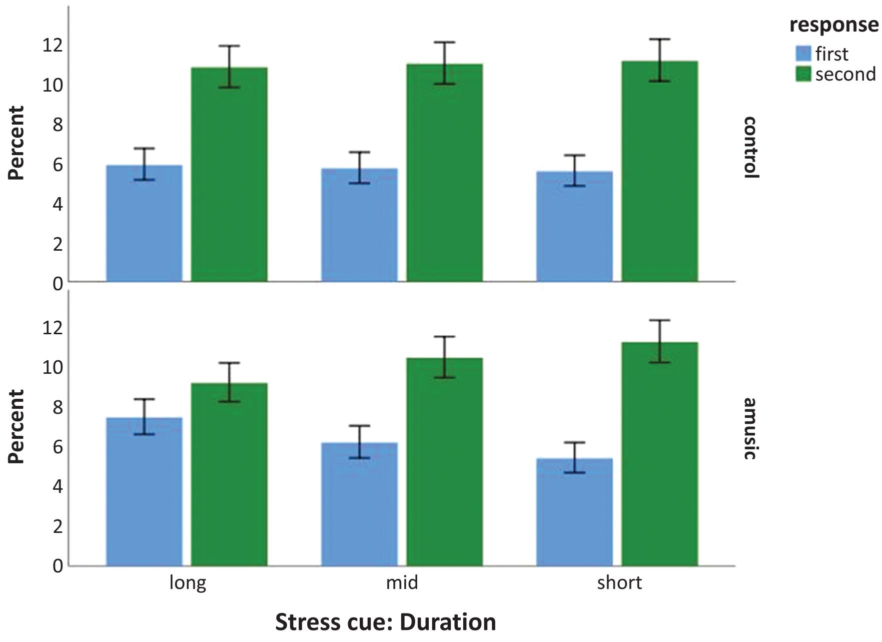

Figure 2: Responses to duration cue split by group. Response first (blue) or second word part (green) in percent (grand total) to long (left), mid (middle) or short duration (right) of the first vowel. Top: control group, bottom: amusic group. Error bars show 95% CI.

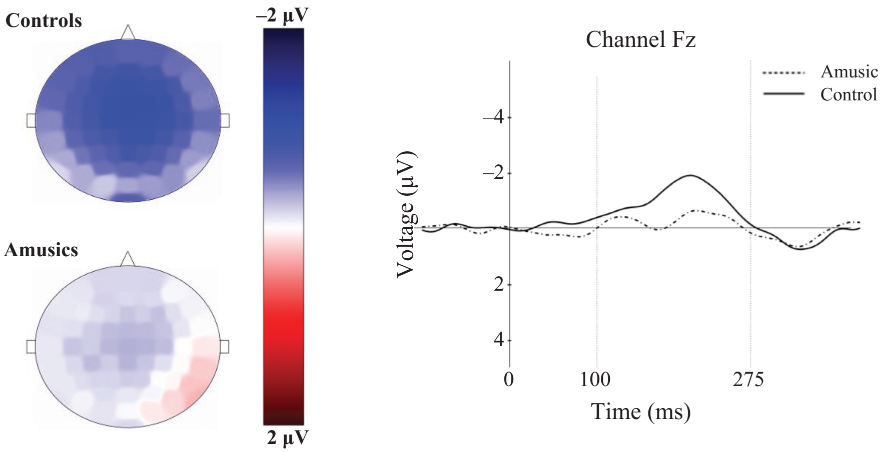

Figure 3: Difference between amusics and controls in a time window between 100 and 275 ms averaged across all conditions. Left panel are the topographical maps and right panel the grand average difference waves plotted at Fz.

deviant minus the waveform of the standard) averaged across all conditions for amusics (dotted line) and controls (solid line).

Figure 4 shows the topographical plot per condition and group. As explained in section 2.1 and 2.3.2, the standards are the two stimuli in line with the natural distribution of cues, with all cues indicating stress either on the first word part (upper two panels; *umstellen*) or on the second word part (lower two panels; *umstellen*). Deviant 1 is the stimulus that was standard in the other condition, i.e. with all congruent stress cues for the opposite stress pattern. Deviant 2 is the stimulus with an ambiguous pitch but all other cues for stress on the first word part. Deviant 3 has a pitch pattern that indicates stress on the first word part but all other cues are in line with stress on the second word part. The dark blue color for the topographical plot on the upper left indicates that the deviant with all cues for stress on the second word part and 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

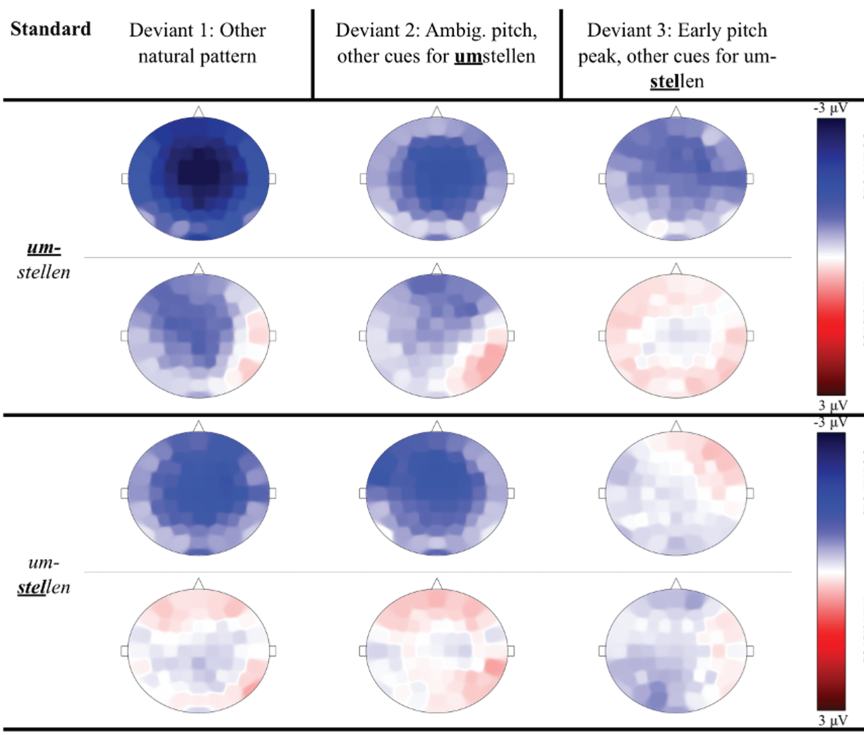

Figure 4: Topographical maps of amusics and controls averaged in a time window between 100 and 275 ms per condition. Controls at the top and amusics at the bottom per condition.

column. The reverse condition, with deviant stressed on the first and standard on the second word part, elicited quite a strong MMN again in controls, but not so in amusics (cf. plot 3 and 4 left column).

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 8.

Table 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: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

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 9: Summary of the model.

	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

A laterality analysis was carried out but revealed no significant differences between conditions or groups.

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 as described in section 3.1:, a summary of the model can be found in Table 9. 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 models with only group as fixed factor were performed for each condition separately. They all yielded significant results, except in one case: When the standard was stressed on the second word part and Deviant 3 from Table 8 was used, then there was no significant effect, also depicted in similar patterns for controls and amusics the bottom right topographical maps in Figure 4.

4 Discussion

Our experiments tested the identification of word stress by amusics and controls based on the acoustic cues of fundamental frequency (pitch), duration, intensity and spectral slope.

Spectral slope and intensity did not have a significant effect on identification. The differences in both cues are based on the naturally produced minimal stress word pair. For spectral slope, it is rather small (a difference of 5). For intensity, the difference is noticeable (8 dB in the first word part and 6 dB in the second). It is possible that taken together the two are still too small and not salient enough to override the more salient acoustic information of duration and pitch contour. Recall also from section 1.4. that loudness and spectral slope, together with pitch, form secondary cues to word stress in German (Lintfert, 2010) while duration has been shown to be the primary perceptual cue (Dogil, 1995; Jessen et al., 1995; Haake et al., 2013).

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 word part (as expected, since default stress is on the penultimate syllable in German; cf. Eisenberg, 1991; Wiese, 2000). 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 word part. 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. The exception is the condition where the standard has stress on the second part, while the deviant had only pitch cues for stress on the first part, while all other cues indicated stress on the second part. For this condition, controls showed the least strong MNN, presumably because the deviant did not differ from the standard with respect to any cues but pitch. Amusics had a very similar MNN for this condition, and therefore the two groups did not show a significant difference.

Despite the difference in amplitude, the MMNs were 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 findings of no difference in the cue usage between amusics and controls in consonants (VOT) 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 Goulet 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 can be criticized for its small sample size, which can be seen as problematic for an ERP study. A small sample size is a pitfall that all studies with amusics face, as the population is a rather small one. Compared to other ERP studies with amusics, however, the size of our sample is fairly large.

5 Conclusion

Our study is the first to investigate word stress processing in congenital amusia. By employing a behavioral and an electrophysiological paradigm, we were able to assess both in relation to each other. Our results surprisingly show that in the behavioral task 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 some difficulties when it comes to word stress cued by pitch, but that these difficulties do not lead to overall processing differences between amusics and controls: While the MMN results indicate deficits in central auditory processing and/or representations, the behavioural results demonstrate that amusics use other cues than pitch and a different weighting of cues than we saw in the control group to compensate for their deficits.

The difficulties of amusics in the perception of word stress that we found in the present study are similar to their previously reported impairment in musical pitch, supporting the accumulating evidence for an impairment that is not specific to the domain of music.

Further studies investigating later ERP components, as well as different cues and cue weightings, and a comparison to other developmental disorders are warranted.

References

Alain, Claude, David L. Woods & Robert T. Knight. 1998. A distributed cortical network for auditory sensory memory in humans. *Brain Research* 812(1–2). 23–37.

Albouy, Philippe, Marion Cousineau, Anne Caclin, Barbara Tillmann & Isabelle Peretz. 2016. Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Scientific Reports* 6. 18861.

Albouy, Philippe, Jérémie Mattout, Romain Bouet, Emmanuel Maby, Gaëtan Sanchez, Pierre-Emmanuel Aguera, Sébastien Daligault, Claude Delpuech, Olivier Bertrand, Anne Caclin & Barbara Tillmann. 2013a. Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain* 136(5). 1639–1661.

Albouy, Philippe, Katrin Schulze, Anne Caclin & Barbara Tillmann. 2013b. Does tonality boost short-term memory in congenital amusia? *Brain Research* 1537. 224–232.
Ayotte, Julie, Isabelle Peretz & Krista Hyde. 2002. Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* 125(2). 238–251.
Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.
Boersma, Paul & David Weenink. 2016. Praat: Doing phonetics by computer (version 6.0.15). Retrieved from http://www.praat.org/.
Braun, Allen, Joe McArdle, Jennifer L. Jones, Vladimir Nechaev, Christopher Zalewski, Carmen Brewer & Dennis Drayna. 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, Jian & Jie Yuan. 2016. The neural causes of congenital amusia. *The Journal of Neuroscience* 36(30). 7803–7804.
Chen, Joyce L., Sukhbinder Kumar, Victoria Wiliamson, Jan Scholz, Timothy Grifiths & Lauren Stewart. 2014. Detection of the arcuate fasciculus in congenital amusia is dependent on tractography algorithm. *The 20th Annual Meeting of the Organization for Human Brain Mapping* 6. 1–11.
Chládková, Katerina, Paola Escudero & Silvia C. Lipski. 2013. Pre-attentive sensitivity to vowel duration reveals native phonology and predicts learning of second-language sounds. *Brain and Language* 126(3). 243–252.
Dalla Bella, Simone, Magdalena Berkowska & Jakub Sowiński. 2011. Disorders of pitch production in tone deafness. *Frontiers in Psychology* 2. 164.
De Bree, Elise, Frank Wijnen & Wim Zonneveld. 2006. Word stress production in three- year-old children at risk of dyslexia. *Journal of Research in Reading* 29(3). 304–317.
Dogil, Gregorz. 1995. The phonetic manifestation of stress. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Universität Stuttgart)* 2. 3–51.
Eisenberg, Peter. 1991. Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. *Zeitschrift für Sprachwissenschaft* 10(1). 37–64.
Fikkert, Paula & Zvi Penner. 1998. Stagnation in prosodic development of language-disordered children. Proceedings of the Boston University Conference on Language Acquisition; p 201–212.
Foxton, Jessica M., Jennifer L. Dean, Rosemary Gee, Isabelle Peretz & Timothy D. Griffiths. 2004. Characterization of deficits in pitch perception underlying “tone deafness”. *Brain* 127(4). 801–810.
Foxton, Jessica M., Rachel K. Nandy & Timothy D. Griffiths. 2006. Rhythm deficits in ‘tone deafness’. *Brain and Cognition* 62(1). 24–29.
Gallon, Nichola, John Harris & Heather Van der Lely. 2007. Non-word repetition: An investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics & Phonetics* 21(6). 435–455.
Gosselin, Nathalie, Pierre Jolicœur & Isabelle Peretz. 2009. Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences* 1169(1). 270–272.
Goswami, Usha, Natasha Mead, Tim Fosker, Martina Huss, Lisa Barnes & Leong Victoria. 2013. Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language* 69(1). 1–17.
Green, David M. & John A. Swets. 1966. *Signal Detection Theory and Psychophysics*. New York: Wiley.

Haake, Caroline, Malte Kob, Klaus Willmes & Frank Domahs. 2013. Word stress processing in specific language impairment: Auditory or representational deficits? *Clinical Linguistics & Phonetics* 27(8). 594–615.
Hamann, Silke, Mats Exter, Jasmin Pfeifer & Marion Krause-Burmester. 2012. Perceiving differences in linguistic and non-linguistic pitch: A pilot study with German congenital amusics. In: Emilios Cambouropoulus, Costas Tsougras, Panayotis Mavromatis, Konstantinos Pastiadis, editors. 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)*.); Thessaloniki, Greece. School of Music Studies, Aristotle University of Thessaloniki. 398–405.
Hay, Rachel A., Brian J. Roach, Vinod H. Srihari, Scott W. Woods, Judith M. Ford & Daniel H. Mathalon. 2015. Equivalent mismatch negativity deficits across deviant types in early illness schizophrenia-spectrum patients. *Biological Psychology* 105. 130–137.
Henry, Molly J. & J. Devin McAuley. 2010. On the prevalence of congenital amusia. *Music Perception* 27(5). 413–418.
Henry, Molly J. & J. Devin McAuley. 2013. Failure to apply signal detection theory to the montreal battery of evaluation of amusia may misdiagnose amusia. *Music Perception* 30 (5). 480–496.
Huang, Wan-Ting, Chang Liu, Qi Dong Dong & Yun Nan. 2015. Categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Frontiers in Psychology* 6. 829.
Huang, Xunan, Caicai Zhang, Feng Shi, Nan Yan & Lan Wang. 2016. Impaired vowel discrimination in Mandarin-speaking congenital amusics. Proceedings of The 5th International Symposium on Tonal Aspects of Languages (Tal 2016); Buffalo, NY. 138–141.
Hutchins, Sean, Nathalie Gosselin & Isabelle Peretz. 2010. Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology* 1. 236.
Hyde, Krista L., Jason P. Lerch, Robert J. Zatorre, Timothy D. Griffiths, Alan Evans & Isabelle Peretz. 2007. Cortical thickness in congenital amusia: When less is better than more. *The Journal of Neuroscience* 27(47). 13028–13032.
Hyde, Krista & Isabelle Peretz. 2004. Brains that are out of tune but in time. *Psychological Science* 15(5). 356–360.
Hyde, Krista, Robert J. Zatorre, Timothy D. Griffiths, Jason P. Lerch & Isabelle Peretz. 2006. Morphometry of the amusic brain: A two-site study. *Brain* 129(10). 2562–2570.
Hyde, Krista, Robert J. Zatorre & Isabelle Peretz. 2011. Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex* 21(2). 292–299.
Jasmin, Kyle, Fred Dick, Lori L. Holt & Adam Tierney. 2019. Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General* 149(5). 914–934.
Jessen, Michael, Krzysztof Marasek, Katrin Schneider & Kathrin Claßen. 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; Stockholm University, Stockholm. 428–431.
Kang, Shinae, Keith Johnson & Gregory Finley. 2016. Effects of native language on compensation for coarticulation. *Speech Communication* 77. 84–100.

Kirmse, Ursula, Sari Ylinen, Mari Tervaniemi, Martti Vainio, Erich Schröger & Thomas Jacobsen. 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.
Kujala, Teija & Risto Näätänen. 2001. The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neuroscience & Biobehavioral Reviews* 25(6). 535–543.
Launay, Jacques, Manon Grube & Lauren Stewart. 2014. Dysrhythmia: A specific congenital rhythm perception deficit. *Frontiers in Psychology* 5. 18.
Lebrun, Marie-Andrée, Patricia Moreau, Andréane McNally, Geneviève Mignault Goulet & Isabelle Peretz. 2012. Congenital amusia in childhood: A case study. *Cortex* 48(6). 683–688.
Leong, Victoria, Jarmo Hämäläinen, Fruzsina Soltész & Usha Goswami. 2011. Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language* 64(1). 59–73.
Lintfert, Britta. 2010. Phonetic and phonological development of stress in German [Doctoral Dissertation]. Stuttgart: University of Stuttgart.
Lisker, Leigh. 1986. “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29(1). 3–11.
Liu, Fang, Cunmei Jiang, Bei Wang, Yi Xu & Aniruddh D. Patel. 2015a. A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia* 66. 111–118.
Liu, Fang, Akshay R. Maggu, Joseph C. Y. Lau & Patrick C. M. Wong. 2015b. Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers. *Frontiers in Human Neuroscience* 8. 1029.
Liu, Fang, Aniruddh D. Patel, Adrian Fourcin & Lauren Stewart. 2010. Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain* 133(6). 1682–1693.
Liu, Fang, Yi Xu, Aniruddh D. Patel, Tom Francart & Cunmei Jiang. 2012. Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition* 79(3). 209–215.
Lolli, Sydney L., Ari D. Lewenstein, Julian Basurto, Sean Winnik & Psyche Loui. 2015. Sound frequency affects speech emotion perception: Results from congenital amusia. *Frontiers in Psychology* 6. 1340.
Loui, Psyche & Gottfried Schlaug. 2009. Investigating musical disorders with diffusion tensor imaging: A comparison of imaging parameters. *Annals of the New York Academy of Sciences* 1169(1). 121–125.
Mandell, Jake, Katrin Schulze & Gottfried Schlaug. 2007. Congenital amusia: An auditory-motor feedback disorder? *Restorative Neurology and Neuroscience* 25(3–4). 323–334.
Marin, Manuela M., Bruno Gingras & Lauren Stewart. 2012a. Perception of musical timbre in congenital amusia: Categorization, discrimination and short-term memory. *Neuropsychologia* 50(3). 367–378.
Marin, Manuela M., William F. Thompson & L. Stewart. 2012b. Emotion perception of dyads and triads in congenital amusia. In: Emilios Cambouropoulus, Costas Tsougras, Panayotis Mavromatis, Konstantinos Pastiadis, editors. 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).*); Thessaloniki, Greece. School of Music Studies, Aristotle University of Thessaloniki. 652–653.

Mignault Goulet, Geneviève, Patricia Moreau, Nicolas Robitaille & Isabelle Peretz. 2012. Congenital amusia persists in the developing brain after daily music listening. *Plos One* 7(5). e36860.

Moreau, Patricia, Pierre Jolicœur & Isabelle Peretz. 2009. Automatic brain responses to pitch changes in congenital amusia. *Annals of the New York Academy of Sciences* 1169(1). 191–194.

Moreau, Patricia, Pierre Jolicœur & Isabelle Peretz. 2013. Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition* 81(3). 337–344.

Näätänen, Risto. 2001. The perception of speech sounds by the human brain as reflected by the Mismatch Negativity (MMN) and its magnetic equivalent (mMMN). *Psychophysiology* 38(1). 1–21.

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

Näätänen, Risto, Pekka Paavilainen, Tehu Rinne & Kimmo Alho. 2007. The Mismatch Negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology* 118(12). 2544–2590.

Näätänen, Risto, Elyse Sussman, Dean Salisbury & Valerie L. Shafer. 2014. Mismatch Negativity (MMN) as an index of cognitive dysfunction. *Brain Topography* 27(4). 451–466.

Nan, Yun, Wan-ting Huang, Wen-jing Wang, Chang Liu & Qi Dong. 2016. Subgroup differences in the lexical tone Mismatch Negativity (MMN) among Mandarin speakers with congenital amusia. *Biological Psychology* 113. 59–67.

Partanen, Eino, Martti Vainio, Teija Kujala & Minna Huotilainen. 2011. Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination. *Psychophysiology* 48(10). 1372–1380.

Patel, Aniruddh D. 2003. Rhythm in language and music: Parallels and differences. *Annals of the New York Academy of Sciences* 999(1). 140–143.

Patel, Aniruddh D., Meredith Wong, Jessica Foxton, Aliette Lochy & Isabelle Peretz. 2008. Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception* 25(4). 357–368.

Peretz, Isabelle, Julie Ayotte, Robert J. Zatorre, Jacques Mehler, Pierre Ahad, Virginia B. Penhune & Benoît Jutras. 2002. Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron* 33(2). 185–191.

Peretz, Isabelle, Anne J. Blood, Virginia B. Penhune & Robert J. Zatorre. 2001. Cortical deafness to dissonance. *Brain* 124(5). 928–940.

Peretz, Isabelle, Anne Sophie Champod & Krista Hyde. 2003. Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences* 999(1). 58–75.

Peretz, Isabelle, Jenny Saffran, Daniele Schön & Nathalie Gosselin. 2012. Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences* 1252(1). 361–366.

Pettinato, Michèle & Jo Verhoeven. 2008. Production and perception of word stress in children and adolescents with Down syndrome. *Down Syndrome Research and Practice* 13. 48–61.

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

Phillips-Silver, Jessica, Petri Toiviainen, Nathalie Gosselin & Isabelle Peretz. 2013. Amusic does not mean unmusical: Beat perception and synchronization ability despite pitch deafness. *Cognitive Neuropsychology* 30(5). 311–331.
Picton, Terence W., Claude Alain, Leun Otten, Walter Ritter & André Achim. 2000. Mismatch Negativity: Different water in the same river. *Audiology and Neurotology* 5(3–4). 111–139.
Repp, Bruno H. & Yi-Huang Su. 2013. Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review* 20(3). 403–452.
Shinozaki, Jun, Nobuo Hiroe, Masa-aki Sato, Takashi Nagamine & Kaoru Sekiyama. 2016. Impact of language on functional connectivity for audiovisual speech integration. *Scientific Reports* 6(1). 31388.
Shtyrov, Yury, Teija Kujala, Satu Palva, Risto J. Ilmoniemi & Risto Näätänen. 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, Alexander, Paavo Alku & Teija Kujala. 2010. Change and novelty detection in speech and non-speech sound streams. *Brain Research* 1327. 77–90.
Sowiński, Jakub & Simone Dalla Bella. 2013. Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia* 51(10). 1952–1963.
Stewart, Lauren. 2008. Fractionating the musical mind: Insights from congenital amusia. *Current Opinion in Neurobiology* 18(2). 127–130.
Summerfield, Quentin. 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7(5). 1074–1095.
Tal, Idan, Edward W. Large, Eshed Rabinovitch, Yi Wei, Charles E. Schroeder, David Poeppel & Elana Zion Golumbic. 2017. Neural entrainment to the beat: The “missing-pulse” phenomenon. *The Journal of Neuroscience* 37(26). 6331–6341.
Tang, Wei, Xi-jian Wang, Jia-qi Li, Chang Liu, Qi Dong & Yun Nan. 2018. Vowel and tone recognition in quiet and in noise among Mandarin-speaking amusics. *Hearing Research* 363. 62–69.
Thompson, William F., Manuela M. Marin & Lauren Stewart. 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 109(46). 19027–19032.
Tillmann, Barbara, Denis Burnham, Sebastien Nguyen, Nicolas Grimault, Nathalie Gosselin & Isabelle Peretz. 2011. Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in Psychology* 2. 120.
Tillmann, Barbara, Philippe Lalitte, Philippe Albouy, Anne Caclin & Emmanuel Bigand. 2016. Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia* 85. 10–18.
Tillmann, Barbara, Katrin Schulze & Jessica M. Foxton. 2009. Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition* 71(3). 259–264.
Weber, Christiane, Anja Hahne, Manuela Friedrich & Angela D. Friederici. 2004. Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research* 18(2). 149–161.
Weber, Christiane, Anja Hahne, Manuela Friedrich & Angela D. Friederici. 2005. Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiologial evidence. *Cognitive Brain Research* 25(1). 180–187.

Wiese, Richard. 2000. *The phonology of German*. Oxford: Oxford University Press.
Williamson, Victoria J., Claire McDonald, Diana Deutsch, Timothy D. Griffiths & Lauren Stewart. 2010. Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology* 6(1). 15–22.
Williamson, Victoria Jane & Lauren Stewart. 2010. Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory* 18(6). 657–669.
Wise, Karen. 2009. Understanding "tone deafness": A multi-componential analysis of perception, cognition, singing and self-perception in adults reporting musical difficulties [Doctoral Dissertation]: Keele University.
Ylinen, Sari, Anna Shestakova, Minna Huotilainen, Paavo Alku & Risto Näätänen. 2006. Mismatch Negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Research* 1072(1). 175–185.
Zhang, Caicai, Jing Shao & Xunan Huang. 2017. Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *Plos One* 12(8).e0183151.