Impaired vowel perception in congenital amusia: Evidence from event-related potentials

Jasmin Pfeifer[1] & Silke Hamann[2]

[1]Institute for English Language and Linguistics, Heinrich-Heine-University, Germany
[2]Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands
[1]pfeifer@phil.hhu.de; [2]silke.hamann@uva.nl

ABSTRACT

Congenital amusia is an auditory perception disorder present from birth. It has been presumed to affect only the musical domain. However, several studies now show that amusics also have impaired perception of intonation and linguistic tones.

We show that amusia also influences linguistically relevant cues other than pitch by investigating the perception of German front high vowels, differentiated by duration and formant frequency. We assessed amusics' electrophysiological responses, more specifically the MMN with a multi-deviant oddball paradigm.

We used an lmer model for the MMN data, finding significant main effects for group with amusics having a smaller MMN than controls.

Our study shows that amusia negatively affects vowel perception and has therefore more far-reaching consequences for speech perception than previously assumed. We found differences in the MMN, reflecting differences in early auditory change detection.

Keywords: Congenital amusia; MMN; vowel perception; duration; formant frequency

1. INTRODUCTION

The condition known as congenital amusia causes lifelong deficits in the ability to perceive pitch and in some cases also rhythm. Insufficient music exposure, hearing loss, or brain damage do not cause the disorder [1]. The most apparent symptoms to the affected individuals are various inabilities in the musical domain such as: Recognition of familiar melodies, detection of out-of-tune notes or singing, or clapping and singing along. Hence, congenital amusia has long been characterized as a music-specific pitch perception disorder [1]. Musical impairments were the focus of initial research on amusia. More recent work, investigated the impact of amusia on language perception, as pitch is also an important cue in language, e.g., to disambiguate questions from statements or to mark focus. Intonation perception was the first area of speech perception shown to be affected by amusia [2, 3]. Pitch is also used in other linguistic areas, which have been shown to be affected as well, such as tone language perception [3]. Due to these findings, congenital amusia is now seen as a domain-general disorder that negatively affects pitch processing in general [4, 5].

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. The quality of a vowel is usually determined by its lowest two formants, F1 and F2. According to the tongue position, vowels are characterized as high vs. low (F1 values), and front vs. back (F2 values). Other aspects, such as nasality, rounding of the lips, and steady state vs. inherent movement also influence formant values, but are neglected in the present study.

First reports on vowel perception by congenital amusics appeared recently, all testing native speakers of tone languages. Huang et al. [6] 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. [5] 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.

These studies considered the differences in vowel quality only and neglected durational differences. 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 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 electro-physiological 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 (for reviews, see [7, 8]). The MMN can be elicited by a change of frequency, duration, intensity, location or pattern in a stream of speech or non-speech sounds [9, 10]. Generally, so-called oddball paradigms are used: A repetitive standard stimulus is presented many times in a row, occasionally interrupted by a deviant differing in an acoustic feature [7, 8]. In so-called multi-deviant oddball paradigms multiple, different deviants within a string of one standard are used. The MMN typically peaks at around 100 to 250 ms after a deviant is presented, if the participants' auditory system has formed a representation of the repetitive aspect of the standard stimulus [7].

The MMN is ideally suited as a starting point in investigating the electrophysiology of vowel discrimination in amusia, as MMN paradigms have long been used in general auditory but also speech perception research [7, 9, 10, 11, 12]. 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. [7, 9]. [9] 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 [11] or both together [9, 12, 13] have been researched with the MMN. Vowel quantity, contrary to other linguistic features, has been shown to be more right-lateralized [9, 12] and to elicit bigger MMN amplitudes than vowel quality changes [9], while simultaneous changes in quantity and quality elicited the biggest amplitude [13]. The MMN thus seems well suited to investigate the neuro-physiological processes underlying congenital amusia.

A number of studies investigating pitch perception in amusia have made use of the MMN already, but with heterogeneous findings: Braun et al. [14] found the MMN to be absent in amusics for melodies containing altered notes, while Moreau et al. [15, 16], using melodies and piano tone sequences, found normal MMNs in amusics. Reduced, abnormal MMNs were found by Nan et al. [17] for a subgroup of tone-language speaking amusics as responses to lexical tones. Taken together, the aforementioned findings show that at least some amusics seem to have 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 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 show reduced MMNs in comparison to controls. We also assess whether the expected deficit is present both in vowel quality and quantity, expecting both will be impaired due to the participants' deficit in pitch and rhythm perception.

2. MATERIALS AND METHODS

2.1. Participants

11 amusics and 11 controls matched for age, gender, handedness, education and musical training participated. All participants were native speakers of German, right-handed and had no self-reported psychological or neurological disorders. They all had normal hearing defined as a mean hearing level of 20 dB or less in both ears. Congenital amusia was diagnosed based on the three pitch-based and the rhythm subtest of the Montreal Battery of Evaluation of Amusia [18] 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.

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, 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 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 [19], see the contrast *Betten* /ɛ/ 'beds' – *bäten* /ɛː/ 'if they requested' – *beten* /eː/ 'to pray', while /e/ occurs only in loanwords in unstressed position, e.g. in *Chemie* 'chemistry'. However, /e/ is considered to be a phoneme by some phonologists [20, 21].

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 1. They are based on acoustic measurements by Jessen [22]. The values of the other two vowels can be extrapolated from this.

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

Table 1: Duration and formant values of the two German vowels /eː/ and /ɛ/ from Jessen [22].

2.3. EEG paradigm, recording and pre-processing

Participants completed a passive listening task while watching a silenced nature documentary without subtitles. 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.

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 vowel was the standard, while the other three vowels served as deviants. The standard occurred 85% of the time and each deviant occurred 5 % of the time. The ISI was varied randomly between 400 ms and 600 ms to avoid entrainment effects to the stimulus chain. 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 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.

The EEG was recorded using a BioSemi Active Two system with 64 active electrodes. 7 further electrodes were placed on the tip of the nose, on the mastoid, below and around the eyes. The EEG signal was recorded at 8192 Hz and later down-sampled to 512 Hz. The subsequent analyses were performed in Praat [23]. The data were offline referenced to the average mastoids channels and bandpass filtered from 1 Hz to 25 Hz. Continuous data was segmented into 500-ms epochs. The EEG was segmented into epochs of 500 ms with a 100-ms baseline. Artifact correction was done automatically and all epochs with activity exceeding +/−75 µV were excluded.

3. RESULTS

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

Visual analysis confirmed the negative polarity, the expected latency and fronto-central scalp distribution of the MMN. Figure 1 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 confirms a difference between amusics and controls.

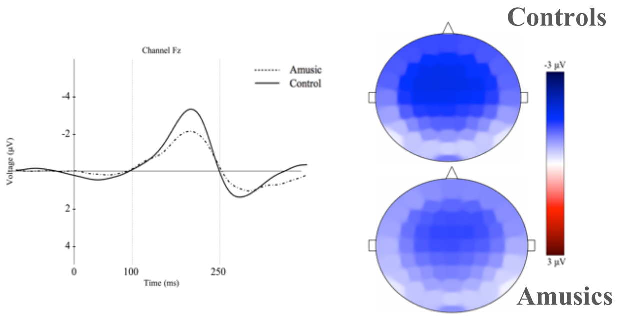

Figure 1: Difference curves (deviant – standard) for amusics and controls in a time window between 100 and 250 ms averaged across all conditions. On the left are the grand average difference waves plotted at Fz, and on the right topographical maps.

Figure 2 shows the grand average difference waves between 100 and 250 ms plotted at Fz for amusics and controls split per condition. The figure illustrates the, on average, greater negativity in the control group, depicted by the solid line. It also seems to indicate differences between the conditions as seen for example in the higher amplitude in the last column.

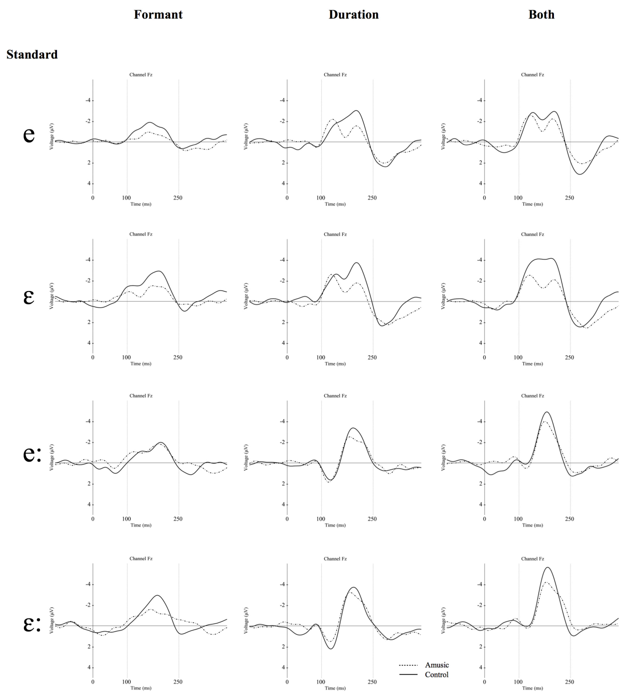

Figure 2: Grand average difference waves plotted at Fz for amusics and controls per condition.

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. Significance levels and mean values at Fz are given in Table 2.

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

Table 2: 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 [24] 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). 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 ($t(2351.8) = 3.85$, $p < 0.001$). Figure 2 serves as a visualisation of this.

To understand the Group by Condition interaction, *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, namely from standard /ɛ:/ to deviant /e:/.

4. DISCUSSION AND CONCLUSION

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 [25]. The general finding that amusics did indeed display an MMN response is in direct opposition to that by Braun et al. [14], 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. [15, 16] who found amusics displayed completely normal MMNs. All these findings utilized musical stimuli, however. Our study supports Nan et al.'s [17] results, who also found a reduced MMN for tone-language speaking amusics in response to lexical tones. Our findings are also in line with [9, 13], as durational changes elicited bigger MMN amplitudes than vowel quality changes, and simultaneous changes in quantity and quality elicited the biggest amplitude.

A point of criticism concerning our 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.

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 concerning the choice of stimuli: We opted for synthesized vowels instead of natural ones as we favored to have control over all parameters, while we ensured at the same time that the vowels sounded as natural as possible. Secondly, we could have chosen four vowels that are all considered German native phonemes by traditional accounts of German phonology such as [19]. The choice of the presently employed mid vowels, however, avoided periphery effects.

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.

We were 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.

Our 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.

5. REFERENCES

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

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

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

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

[5] 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.

[6] 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).

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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. *PLoSONE, 3*(6), 1–6.

[15] 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*(1), 191–194.

[16] 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.

[17] 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. *Biol Psychol, 113*, 59–67.

[18] 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.

[19] Wiese, R. 1996. *The Phonology of German.* Oxford: Clarendon Press.

[20] Wurzel, W. U. 1981. Phonologie: Segmentale Struktur. In: K. E. Heidolph et al. (ed.) *Grundzüge einer deutschen Grammatik.* 898–990. Berlin: Akademie-Verlag.

[21] Giegerich, H. 1985. *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press.

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

[23] Boersma, P., Weenink, D. 2016. *Praat: doing phonetics by computer* (Version 6.0.15). Retrieved from http://www.praat.org/

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

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