

Perceiving Differences in Linguistic and Non-Linguistic Pitch: A Pilot Study With German Congenital Amusics

Silke Hamann,^{*1} Mats Exter,^{#2} Jasmin Pfeifer,^{#3} Marion Krause-Burmester^{#4}

^{*}*Amsterdam Centre for Language and Communication, University of Amsterdam, The Netherlands*

[#]*Institute for Language and Information, University of Düsseldorf, Germany*

¹silke.hamann@uva.nl, ²exter@phil.hhu.de, ³pfeifer@phil.hhu.de, ⁴krause@phil.hhu.de

ABSTRACT

This study investigates the perception of pitch differences by seven German congenital amusics in speech and two types of non-speech material (sinusoidal waves and pulse trains). Congenital amusia is defined by a deficit in musical pitch perception, and recent studies indicate that at least a subgroup of congenital amusics also show deficits in linguistic pitch perception.

While previous studies employed pitch differences that occur in naturally spoken pairs of statements vs. echo questions to test the influence of amusia on linguistic pitch perception, the present study parametrically varied the pitch differences in steps of one semitone (from one to seven semitones). We further tested the influence of the direction of the pitch change, the length of the stimuli and the continuity of the pitch curve.

Our results show that amusics have difficulties detecting pitch changes both in non-linguistic stimuli and in speech. Furthermore, we found that amusics and controls performed better when the stimuli were discontinuous and the pitch was raised (instead of lowered). With respect to non-speech material, all participants performed better for pulse trains. The length of the stimuli did not influence the performance of the participants.

I. INTRODUCTION

Congenital amusia is a neuro-developmental disorder that has a negative influence on pitch perception (Peretz et al., 2002; Foxton et al., 2004; Stewart, 2008). People with congenital amusia (in the following called *amusics*) face life-long impairments in the musical domain, and their symptoms can be so severe that music causes discomfort to them (Stewart, 2008, p. 127). Amusics have deficits in fine-grained pitch discrimination, i.e. they cannot detect a pitch difference between two adjacent tones if this difference is one semitone or less (Peretz et al., 2002; Foxton et al., 2004).

Congenital amusia is neither caused by insufficient exposure to music, nor by a hearing deficiency, brain damage or intellectual impairment (Ayotte et al., 2002). The underlying cause of this 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) or a working-memory deficit specific to non-verbal sequences (Gosselin et al., 2009; Tillmann et al., 2009), though conclusive evidence for either of these two hypotheses is still missing.

This condition is particularly interesting to linguists because of the ongoing debate whether language is affected (cf. Lochy et al., 2004; Patel et al., 2008; Hutchins et al., 2010; Liu et al., 2010) or not (cf. Ayotte et al., 2002). It has long been argued that congenital amusia is domain-specific to music and does not affect language (e.g. Peretz et al., 2002; Ayotte et al., 2002). However, recent studies such as Patel et

al. (2008) and Liu et al. (2010) suggest that this view has to be reconsidered and that amusics show deficits in the perception of linguistic pitch (intonation).

Patel et al. (2008) investigated the pitch perception of British English and Canadian French amusics in an AX (“same–different”) discrimination task. The study used natural sentences (statement–question pairs) that were cross-spliced and edited so that they differed acoustically only in the final region of the intonation contour. In addition, they used tonal analogs of the statement–question pairs. Patel et al. found that 30% of the amusics had difficulties discriminating statements from questions based on intonation, while they were able to discriminate the tone analogs modeled on the basis of these sentences quite well (Patel et al., 2008, p. 364). These findings are in contrast to previous studies, such as Ayotte et al. (2002), who found that amusics showed no impairment of linguistic pitch perception.

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

The present pilot study attempts to replicate the findings of these previous studies on Canadian French and British English for German amusics. It examines the discrimination of linguistic pitch and two types of tonal analogs by seven amusics and 35 matched controls. In contrast to the studies by Patel et al. and Liu et al., the present study employs a parametric manipulation of small pitch differences (from one to seven semitones, in steps of one semitone). It further investigates the influence of three parameters on the amusics’ perception of linguistic and non-linguistic pitch, namely the direction of pitch change, the length of stimuli, and the continuity of the pitch curve. A case study by Peretz et al. (2002) indicated that the direction of the pitch change might have an influence on the performance of amusics, as their amusic could only perceive a rising difference. With respect to the length of stimuli, studies proposing a memory deficit (Gosselin et al., 2009; Tillmann et al., 2009) showed that amusics perform worse for long stimuli. No previous studies are known to us that investigated the influence of pitch continuity on the performance of amusics.

Table 1. Characteristics of the participants (all right-handed). Scores for the MBEA subtests are the number of correct responses (with a maximum of 30). Boldface scores are below the cut-off score (see text).

| Participant | Age | Gender | Years of Education | MBEA Scale | MBEA Contour | MBEA Interval | MBEA Rhythm |
|---------------------|---------------|--------|--------------------|--------------|--------------|---------------|--------------|
| Amusic 1 | 19 | female | 13 | 19 | 23 | 20 | 27 |
| Amusic 2 | 23 | male | 18 | 20 | 21 | 23 | 19 |
| Amusic 3 | 30 | female | 22 | 20 | 22 | 18 | 19 |
| Amusic 4 | 35 | female | 18 | 28 | 19 | 22 | 21 |
| Amusic 5 | 23 | female | 14 | 22 | 21 | 21 | 19 |
| Amusic 6 | 22 | female | 16 | 26 | 22 | 23 | 17 |
| Amusic 7 | 51 | female | 10 | 21 | 22 | 23 | 18 |
| Amusic Mean | 29.0 | | 15.9 | 22.3 | 21.4 | 21.4 | 20.0 |
| (SD) | (10.3) | | (3.6) | (3.1) | (1.2) | (1.8) | (3.1) |
| Control Mean | 28.3 | | 16.8 | 27.5 | | | 24.5 |
| (SD) | (9.2) | | (2.7) | (2.9) | | | (3.0) |

II. METHOD

A. Participants

We tested seven amusics and 35 matched controls (all German native speakers). The amusic participants were recruited via advertisement. They were screened with a questionnaire on their experience with music and tested with the *Montreal Battery of Evaluation of Amusia* (MBEA).

The MBEA is a series of tests designed by Peretz et al. (2003) that was originally devised to assess the musical abilities of brain-damaged patients. Nowadays, it is the main tool used to diagnose congenital amusia. It consists of six subtests, namely a scale, contour, interval, rhythm, meter, and memory test. A score of 22 or below (out of 30) on at least two of the first four subtests was used to diagnose congenital amusics in the present study (a similar procedure was employed by Liu et al. 2010 and Williamson et al. 2010, see Section IV-B for a discussion). A cut-off score of 22 was calculated by Peretz et al. (2003) based on the MBEA scores for 160 normal participants and the assumption that the performance of amusics lies at least 2 standard deviations below the mean scores.

The control group was matched for handedness, age, gender (five males), and years of education; see participant details in Table 1. Controls were assessed with a shortened version of the MBEA consisting only of the scale and rhythm subtests (Peretz et al., 2003) to exclude possible deficits in the musical domain.

All participants were free of neurological and psychiatric disorders. They had normal hearing (defined as a mean hearing level of 20 dB or less in both ears), which was assessed by pure tone audiometry at 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz preceding the experiment.

B. Stimuli

A male native speaker of German was recorded in a sound-attenuated booth with a Sennheiser ME 62 microphone and a Sound Devices MixPre microphone preamplifier/mixer onto a Marantz PMD570 solid state recorder with a sampling frequency of 44.1 kHz. The speaker read a story that contained eight target sentences. Half of these sentences were

statements, the other half echo questions that differed from the statements only in the final pitch direction (rising instead of falling). They further differed in length, i.e. they were either short, as in (1a) and (2a), with 3 or 6 syllables and a mean duration of 1.05 s, or long, as in (1b) and (2b), with 7 or 10 syllables and a mean duration of 1.59 s.

Thirdly, the sentences either consisted only of voiced sonorants (i.e., segments having a periodic, well-defined formant structure (Ladefoged 1997, p.615)), cf. (1), or they also included voiceless obstruents (i.e., segments that lack a formant structure), such as the plosive [p], the affricate [tʃ] and the fricative [f], cf. (2). Sentences with only sonorants result in a continuous pitch contour, while sentences which also include obstruents result in a discontinuous pitch contour (cf. Figure 1).

- (1) a. Jan nahm Lamm.
'Jan chose lamb.'
- b. Jan nahm lange Jahre Lamm.
'Jan had chosen lamb for years.'
- (2) a. Atze hatte Pappe.
'Atze had cardboard.'
- b. Atze hatte anfangs schwarze Pappe.
'In the beginning, Atze had black cardboard.'

The male speaker read the eight target sentences with a mean fundamental frequency of 106.6 Hz. Using Praat (Boersma & Weenink, 2011), the pitch contour of each target sentence was extracted, stylized and simplified in such a way that every pitch point that lay closer than two semitones to its immediate neighbor(s) was deleted. This simplified pitch contour was then used to replace the pitch contour of the target sentence, resulting in a synthesized stimulus. For each simplified pitch contour, seven further pitch contours were created by moving the final pitch point in seven one-semitone steps either upward (for statement sentences) or downward (for question sentences). Each of the resulting seven pitch contours was used to replace the original contour, yielding a set of eight synthesized stimuli for each target sentence, and a total of 64 stimuli. Illustrations for the synthesized pitch contours based on the target sentences (1a) and (2b) are given in Figure 1.

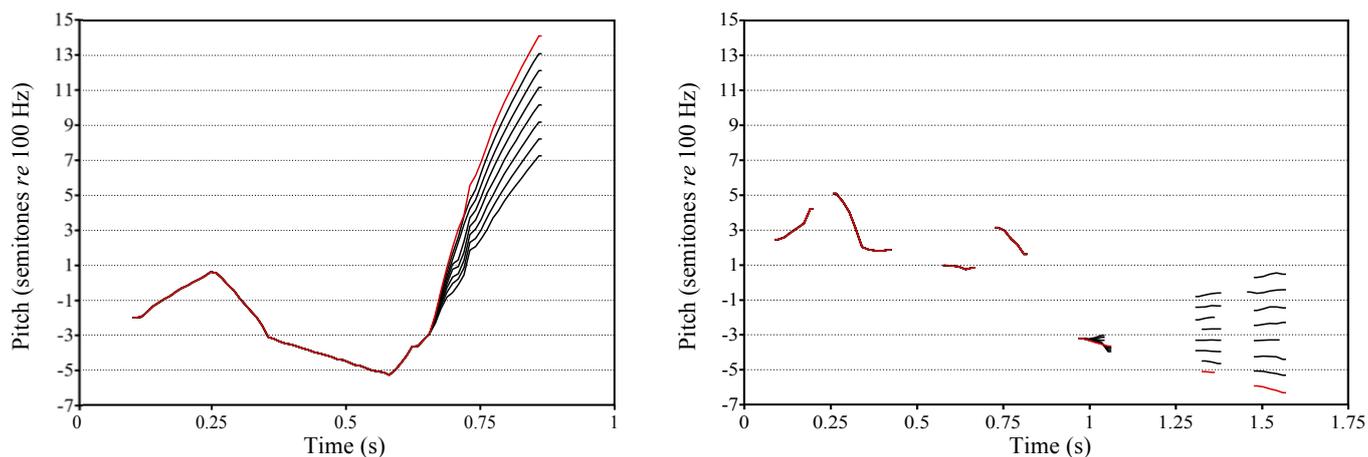


Figure 1. The sets of pitch contours created from the sentences *Jan nahm Lamm?* (short, continuous, downward) on the left and *Atze hatte anfangs schwarze Pappe* (long, discontinuous, upward) on the right, with the original simplified pitch contours (in red) and the seven manipulations (in black).

In addition to the speech stimuli, two types of non-speech stimuli, namely sinusoidal wave analogs and pulse train analogs, were created. For the sinusoidal wave analogs, each pitch contour created for the synthesized sentences described above was converted into a sinusoidal wave whose frequency followed that of the simplified pitch contour. Onsets and offsets of the sinusoidal waves were filtered with a low-pass filter of 200 Hz to remove unwanted transients. The pulse train analogs were created by turning the sine analogs into sequences of pulses, with the distance between pulses inversely proportional to the frequency of the pitch curve (a higher frequency corresponding to a smaller distance). In contrast to the pure sinusoidal tones, the pulse train analogs have a more or less flat spectrum. The two additional non-speech stimuli for each synthesized sentence resulted in a total of 192 stimuli.

The stimuli were paired within each set of eight synthesized stimuli created from the same target sentence. Each stimulus created by raising or lowering the final pitch point was once paired with itself and once with the simplified pitch stimulus it was created from. This resulted in 14 stimuli pairs per set (an equal number of *same* and *different* pairs), and a total of 336 experimental stimuli pairs.

The 336 experimental stimuli pairs differed in five conditions, cf. (3).

(3) *Experimental conditions:*

- Continuity (discontinuous, continuous)
- Length (short, long)
- Direction (upward, downward)
- Type (speech, sine, pulses)
- Interval (0 to 7 semitones)

In addition to the 336 experimental stimuli pairs, 12 catch trials and nine practice trials were created. For the catch trials, the simplified pitch of the four short target sentences was changed in a step of 24 semitones (upwards or downwards). These pitch contours were then used to synthesize new sentences, which were paired with the simplified sentences

they were created from. This was done for both continuity types and both intonation contours, resulting in 12 catch pairs. These catch trials were included in the experiment to ensure participants paid attention and performed the task correctly. Both controls and amusics perceived all catch pairs as different, and thus no participant had to be excluded on these grounds.

The practice trials were used in the practice session before the experiment and were created from the simplified pitch contour of the short, continuous, upward sentence by moving upwards the final pitch point in steps of 2.5 and 5.5 semitones, respectively. For both resulting pitch contours a sentence, a sinusoidal tone analog and a pulse train analog was created. Each of these six stimuli was paired with the corresponding simplified pitch stimulus that was not manipulated in final pitch, and the latter were also paired with themselves. This resulted in six *different* practice trials and three *same* practice trials.

C. Design and Procedure

For the perception experiment, an AX (“same–different”) discrimination task was used. The experiment took place in the phonetics laboratory at the University of Düsseldorf and lasted approximately 60 minutes. Participants were seated in a sound-attenuated booth, and the stimuli were presented over AKG K 601 headphones using Praat on a Windows XP computer. Participants could adjust the volume to a level they felt comfortable with. They were asked to listen carefully to each stimulus pair and to decide whether the two stimuli were the same or different. They were told to respond as quickly as possible by pressing labeled buttons on the keyboard in front of them.

Each stimulus pair was presented in the same way: A warning signal was followed by one second of silence. Then the stimulus pair was played, with an inter-stimulus interval (ISI) of one second. This ISI was chosen since longer ISIs can cause an interference with possible pitch memory deficits in amusics (Williamson et al., 2010). At the same time, an ISI of one second is considered long enough not to risk problems

with rapid auditory processing that might also be problematic for amusics (Williamson et al., 2010).

At the beginning of the experiment, participants had a practice session with the nine practice trials to familiarize themselves with the experimental procedure and the different stimulus types (speech, sine, pulses) they would hear. During the practice session, participants received feedback whether their judgment was correct or incorrect. No feedback was provided during the following test phase.

In the test phase, the 336 experimental stimulus pairs were blocked: The 14 pairs (and a possible catch trial) that shared all conditions (continuity, length, direction and type) were presented within one block in randomized order. Across blocks, the order was pseudo-randomized so that two blocks with more than two conditions in common did not immediately follow each other. The order of blocks was counterbalanced across participants to compensate for possible symptoms of fatigue.

After every eighth block a break occurred (two in total) in which participants had to read aloud stories and answer questions concerning these stories.

All participants received a small monetary reimbursement for their participation.

III. RESULTS

To assess the results of the experiments described above, a number of statistical procedures were performed on the data, using the statistical package R (R Development Core Team, 2012).

A. Sensitivity and Perceptual Threshold

First, participants' responses were scored as *hits* when they correctly identified differing stimuli as different, and as *misses* when differing stimuli were identified as same. Conversely, responses were recorded as *false alarms* when participants falsely identified identical stimuli as different, and as *correct rejections* when they correctly identified identical stimuli as same.

Hits (H), misses (M), false alarms (FA), and correct rejections (CR) were counted, and the counts entered into the calculation of three sensitivity measures as follows:

- Hit rate:
 $HR = H/(H + M)$
- Percentage correct:
 $PC = 100 \cdot (H + CR)/(H + M + FA + CR)$
- d' ("d-prime"):
 $d' = Z(HR) - Z(FAR)$,

where $FAR = FA/(FA + CR)$ is the false-alarm rate and $Z(p)$ is the inverse of the standard normal cumulative distribution function (Signal Detection Theory: Green & Swets, 1966; Macmillan & Creelman, 2005).

The three sensitivity measures (hit rate, percentage correct, and d') were calculated for each participant ($N = 42$), pooled over all experimental conditions. We then investigated the extent to which the two groups (experimental and control) differed with regard to these three measures, as well as the statistical significance of those differences.

Firstly, the difference in hit rate between the groups was significant, $t(7.53) = 4.52$, $p = .0023$. For the experimental group, the mean and standard deviation were $M = .48$ and $SD = .17$, respectively; the control group, on the other hand, had a mean $M = .78$ and a standard deviation $SD = .13$. The 95% confidence interval (CI) for the absolute difference between the means was $[.15, .46]$.

Secondly, the groups were also significantly different with regard to percentage correct, $t(8.35) = 4.04$, $p = .0034$ (experimental group: $M = 68.58\%$, $SD = 8.01\%$; control group: $M = 81.86\%$, $SD = 7.66\%$). The absolute difference between the means showed a 95% CI of $[5.75, 20.82]$.

Finally, the third measure, d' , also revealed a significant difference between the two groups, $t(10.42) = 3.57$, $p = .0048$ (experimental group: $M = 1.28$, $SD = 0.46$; control group: $M = 2.00$, $SD = 0.60$), with a 95% CI for the absolute difference between the means of $[0.27, 1.17]$.

It should be noted, however, that (for our data at least) the difference between the experimental and control groups in terms of percentage correct and d' can in turn be attributed to the consistent difference in hit rate, which is implicit in the calculation of the other two measures. No comparably clear difference between the two groups emerged with respect to false-alarm rate, and the two measures (hit rate and false-alarm rate) were only weakly correlated over all participants, $r = 0.22$.

Next, we investigated the question of how the d' measure (as the most appropriate measure of sensitivity) within the two groups as a whole differed for each interval size. As shown in Figure 2, d' was consistently lower at all interval sizes for the experimental group than for the control group.

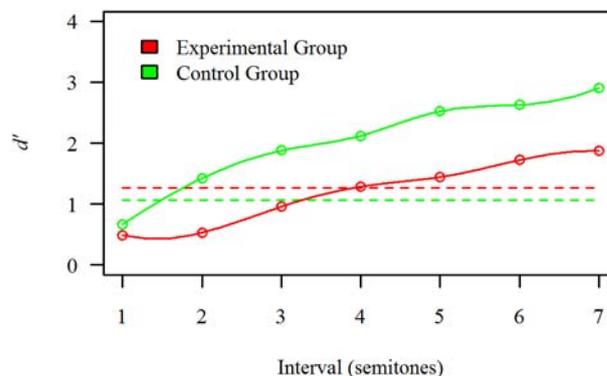


Figure 2. Values of d' of the experimental and control groups for each interval size (from 1 to 7). Individual values are connected by a line representing interpolated values. Dashed lines indicate d' at chance hit rate for the two groups (see text).

As explained above, both hit rate and false-alarm rate enter into the calculation of d' . Now, taking hit rate as a measure of how well an existing difference in the stimuli is actually perceived, and considering that responding to the stimuli purely by chance would result in a hit rate of .5, we argue that the interval size at which the hit rate has a value of .5 can be plausibly interpreted as a kind of perceptual threshold for interval size, at least in the context of our data. False-alarm rate, on the other hand (being calculated for an interval size of 0) is constant *within* each group but different *between* the groups. This results in different d' values at chance hit rate for

the two groups (experimental group: 1.26; control group: 1.07).

To quantify the perceptual threshold in terms of d' , a cubic spline interpolation was performed on the values given in Table 2, resulting in a perceptual threshold of 3.80 semitones for the experimental group and 1.67 semitones for the control group.

Table 2. Values of d' of the experimental and control groups for each interval size (from 1 to 7).

| Interval (semitones) | Experimental Group | Control Group |
|----------------------|--------------------|---------------|
| 1 | 0.49 | 0.66 |
| 2 | 0.53 | 1.42 |
| 3 | 0.96 | 1.88 |
| 4 | 1.29 | 2.12 |
| 5 | 1.44 | 2.53 |
| 6 | 1.72 | 2.63 |
| 7 | 1.88 | 2.90 |

B. Regression Analysis

To assess the influence and statistical significance of the various experimental conditions (cf. Section II-B), a mixed-effects logistic regression analysis was performed on the data ($N = 14,112$ observations). This model takes into account the specific design of the experiment (repeated measures, dichotomous response variable), and it was specified as follows:

1) *1 dichotomous response variable:*

- Response (0 = same, 1 = different)

2) *4 dichotomous predictor variables:*

- Group (0 = experimental, 1 = control)
- Continuity (0 = discontinuous, 1 = continuous)
- Length (0 = short, 1 = long)
- Direction (0 = upward, 1 = downward)

3) *1 discrete predictor variable:*

- Type (0 = pulses, 1 = sine, 2 = speech)

4) *1 continuous predictor variable:*

- Interval (0 to 7 semitones)

5) *1 discrete random-effect variable:*

- Subject (42 participants)

Note that the variable Direction should be interpreted as follows: 0 = statement (with final tone systematically shifted *upward*), 1 = question (with final tone systematically shifted *downward*).

For the analysis, the discrete predictor variable Type was split up into two dichotomous dummy variables, Sine (0 = pulses, 1 = sine) and Speech (0 = pulses, 1 = speech).

As a first step, a maximally specified model (with all possible interactions) was fit to the data. This resulted in an extremely complex pattern of significances on the one hand and a large number of insignificant terms (at the .05 level) on the other hand. We therefore chose to use a more restricted model including only those two-way interaction terms that included the variable Group (experimental vs. control), thus

enabling us to investigate any differences in the effects of the other variables between the two groups.

Fitting this more restricted model to the data, it was revealed that the variable Length did not reach significance (at the .05 level), neither as a main effect nor in the corresponding interaction term. We therefore excluded the variable and the interaction term from the model, resulting in a final model with 6 predictor variables and 5 interaction terms.

To assess the validity of the model as a whole, a likelihood ratio test was performed first, $\chi^2(11) = 7605.84$, $p < .001$, indicating that the model thus specified fits the data significantly better than an intercept-only model (null model).

Next, the influence and significance of the individual predictor variables and interaction terms were investigated; the results are given in Table 3.

Table 3. Results of mixed-effects logistic regression analysis ($N = 14,112$).

| Coefficient | β | SE β | z Value | p | exp(β) |
|------------------------------|---------|------------|---------|--------|----------------|
| Intercept | -1.82 | 0.32 | -5.72 | < .001 | 0.16 |
| Group | 0.46 | 0.35 | 1.32 | 0.19 | 1.59 |
| Continuity | -0.32 | 0.11 | -2.91 | 0.0036 | 0.72 |
| Direction | -0.32 | 0.11 | -2.91 | 0.0036 | 0.72 |
| Sine | -0.46 | 0.14 | -3.33 | < .001 | 0.63 |
| Speech | 0.035 | 0.13 | 0.27 | 0.79 | 1.04 |
| Interval | 0.52 | 0.024 | 22.10 | < .001 | 1.68 |
| Group \times Continuity | -0.12 | 0.12 | -0.94 | 0.35 | 0.89 |
| Group \times Direction | 0.19 | 0.12 | 1.57 | 0.12 | 1.21 |
| Group \times Sine | 0.21 | 0.15 | 1.39 | 0.16 | 1.24 |
| Group \times Speech | -0.39 | 0.15 | -2.63 | 0.0085 | 0.68 |
| Group \times Interval | 0.45 | 0.029 | 15.43 | < .001 | 1.57 |

As can be seen from the respective p values, the main effects Continuity, Direction, Sine, and Interval, as well as the interaction terms Group \times Speech and Group \times Interval are statistically significant (at the .05 level). The exponentiated coefficients for Speech and Interval (representing odds ratios and commonly used as measures of effect size in logistic regression analysis) can therefore be calculated as follows:

1) *Speech:*

- Group = 0:
 $\exp(\beta) = \exp(0.035) = 1.04$
- Group = 1:
 $\exp(\beta) = \exp(0.035 - 0.39) = 0.70$

2) *Interval:*

- Group = 0:
 $\exp(\beta) = \exp(0.52) = 1.68$
- Group = 1:
 $\exp(\beta) = \exp(0.52 + 0.45) = 2.65$

The implications of these results will be discussed in Section IV below.

IV. DISCUSSION

On all three sensitivity measures (hit rate, percentage correct, and d') applied in this study, amusics performed worse than the control group. As the three measures cannot be considered independent of each other, the following discussion is restricted to d' . While controls had a mean d' of 2.00, amusics showed a mean d' of 1.28. This supports the findings of previous studies that amusics have severe difficulties perceiving pitch differences (in speech and non-speech material). The fact that the control group did not perform at ceiling (as evidenced by a mean percentage correct of only 81.86%) indicates that the present experiment involved a difficult task.

In the present study, the perceptual threshold for the amusic group lies at 3.80 semitones, i.e. the amusics can reliably perceive differences in two final pitch contours if they are at least 3.80 semitones apart. For the control group, this threshold lies at 1.67 semitones. Both values are considerably higher than thresholds found in other studies. For non-amusics, a pitch perception threshold around 0.5 semitones has been reported (Peretz et al., 2002; Hyde & Peretz, 2004; Tillmann et al., 2009). For amusics, the threshold in previous studies varies between slightly above 1 semitone (Foxton et al., 2004), 1.32 semitones (Tillmann et al., 2009), and 2 semitones (Peretz et al., 2002; Hyde & Peretz, 2004). The rather high perceptual thresholds in the present study can again be attributed to the difficulty of the task, in particular, the fact that the two pitch targets to be compared were not adjacent to each other. The difference in perception thresholds between the amusics and the non-amusics in our study is comparable to the difference found by Tillmann et al. (2009).

A. Influence of variables

The mixed-effects logistic regression analysis in Section III-B showed a significant interaction of the variables Group and Interval and a main effect of Interval. While amusics performed slightly better (in the sense that the odds of responding *different* rather than *same* increased by a factor of 1.68) when the interval was increased by one semitone, the same change resulted in a much better performance by the controls (by a factor of 2.65). This is in accordance with all previous findings that amusics perform worse than non-amusic in discriminating tone differences (which correlate with interval size).

Furthermore, there was an interaction of Group and Speech. Controls showed a noticeable decrease in performance for speech as compared to pulses (by a factor of 0.70), while the performance of amusics was almost the same for both conditions (there was a slight increase by a factor of 1.04 for speech as compared to pulses). This means that the controls performed *worse* for linguistic stimuli than for both types of non-linguistic stimuli, whereas the amusics performed *better* for linguistic stimuli than for non-linguistic stimuli. For some reason, non-amusics find a difference in linguistic pitch harder to detect than a difference in musical pitch. We can only speculate that the presence of linguistic material draws the non-amusics' attention away from the pitch, while in the non-linguistic stimuli all attention can be given to pitch differences. Since amusics in general perform worse than

non-amusics, the presence or absence of linguistic material seems to make no difference in their (bad) performance.

The significant main effects of the variables Continuity, Direction, and Sine can be interpreted in the following way. For Continuity, we found that both groups showed more difficulties discriminating continuous than discontinuous stimuli. We hypothesize that it is easier to detect a final pitch difference when the listener has discrete stretches of pitch and therefore can focus on the final stretch instead of having to remember the pitch movement of the whole, continuous sequence.

The direction of pitch change influenced the performance of our participants in the following way: They performed better for a change upwards, i.e. if the second stimulus in a pair had a higher final pitch compared to the first one (with a statement intonation). This finding is in line with the amusia study by Peretz et al. (2002), who report that their subject Monica could only detect rising changes. We propose the asymmetrical performance in our study is due to the fact that German listeners perceptually compensate for a decrease in the intonation curve that usually occurs in their native language towards the end of a statement phrase or utterance (Mixdorff & Fujisaki, 1995). For that reason, German listeners are less sensitive to a lowered final pitch. Perceptual compensation for lowering can also explain the behavior of French listeners like Monica, since French also has a default falling intonation for statements (Hirst & Di Cristo, 1984).

With respect to the variable Sine, the pulse trains were better discriminated than the sinusoidal waves. Since no previous study employed pulse trains, our results cannot be compared to earlier findings. We observed, however, that the pulse trains employed in this study gave the auditory impression of being louder than the other two stimulus types (and in particular, louder than the sinusoidal waves), although acoustically they were not. This perceived loudness of the pulse trains might have facilitated their discrimination. Future studies are necessary to test this hypothesis.

The fact that length of the stimuli did not influence the performance of either controls or amusics points in the direction that amusia is not caused by a memory deficit (contra Gosselin et al., 2009; Tillmann et al., 2009).

B. Diagnosing amusia with the MBEA

Previous studies show no homogeneous use of the MBEA but differ in how many subtests they involve in their diagnosis and in the cut-off scores for these subtests.

Ayotte et al. (2002) use all six subtests of the MBEA but report that there is no single subtest that can reliably distinguish amusics from controls. Foxton et al. (2004) also use all six subtests. They found that all of their amusics show impairment on at least three of the six tests. Liu et al. (2010) and Williamson et al. (2010) use only the first four subtests, as in the present study. Reasons for this might be that the last subtest is on melodic memory (thus tests memory rather than musical abilities) and the fifth, metrical subtest is difficult to perform without musical knowledge, as participants are asked to categorize short tonal sequences as waltz or march. Ayotte et al. (2002) found that amusics did not necessarily perform below cut-off score while non-amusic participants performed on average quite poorly for this subtest.

With respect to cut-off scores for the MBEA, most studies use a value of 22 (based on Peretz et al. 2003). However, Ayotte et al. (2002) employ a cut-off score of three standard deviations below the mean of the controls. Tillmann et al. (2009) report a cut-off score of 23 (22 for the scale subtest). Liu et al. (2010) devise a composite score for three of the four MBEA subtests they employ. The problem of the arbitrariness of cut-off scores for the MBEA has already been noted by Henry & McAuley (2010).

A number of studies use other diagnostics in addition to or instead of the MBEA. Thompson (2007) uses two additional tests, namely the *Beat Alignment Test* (Iversen et al. 2008) and the *Macquarie Monotonic Rhythm Test*. Kalmus and Fry (1980) developed and used the *Distorted Tunes Test*, which was later also used by Jones et al. (2009). Ayotte et al. (2002) and Peretz et al. (2003) employ a questionnaire on the musical abilities of their participants in addition to the MBEA for a classification as amusic or non-amusic.

This heterogeneity in the use of the MBEA raises some concerns about its ability as sole diagnostic for amusia. It has to be noted that the MBEA was originally designed to assess the musical abilities of brain-damaged patients who are diagnosed on the basis of neurological findings. Peretz et al. (2003) themselves raise doubts whether this test is fit to diagnose the presence of a disorder by itself. However, all recent studies on amusia (with the exception of Jones et al. 2009) employ the MBEA as diagnostic tool for amusia.

To assess the reliability of the diagnosis *amusia* gained from the application of the MBEA in the present study, a *k*-means cluster analysis with 2 clusters (hypothesized to correspond to amusic vs. non-amusic participants) was performed on the three sensitivity measures (cf. Section III-A). The cluster analysis based on hit rate yielded a categorization that corresponded closely to the MBEA diagnoses, whereas the categorizations by percentage correct and, surprisingly, *d'* did not reveal any such close correlations to the MBEA results.

The two categories based on hit rate are given in Figure 3.

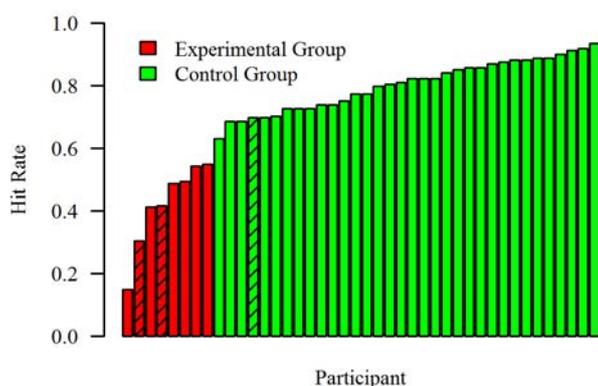


Figure 2. Categorization of participants into the experimental group or control group based on a *k*-means cluster analysis over the hit rate values for each participant. The three hatched bars correspond to the reclassified participants (see text).

The categorization in Figure 3 differs in three respects from the categorization of participants based on the MBEA: One amusic participant (Amusic 6 in Table 1) clustered with the participants of the “control cluster”, and two control

participants clustered with the participants of the “experimental cluster”. This discrepancy between the MBEA classification of the participants and their performance could be an indication for the problematic nature of using the MBEA as sole diagnostic tool.

A further point of discussion with respect to the MBEA is the fact that amusics show differences in scores between subtests that refer to melodic abilities (scale, contour and interval) and those that refer to temporal abilities (rhythm and meter). Peretz et al. (2003) report that one of their 24 tested amusics had scores in the normal range for the melodic subtests, and half of them had high scores in the rhythmical subtests. The amusics in the studies by Ayotte et al. (2002) and Foxton et al. (2004) showed a similar variable performance on the rhythmical test. Peretz et al. conclude that amusia seems to involve a core deficit in pitch perception, but also that different types of amusia seem to exist (Peretz et al. 2003, p. 70).

In our study, all amusics had a score of 23 or below for the interval and contour subtests, thus performed quite badly for those two tests. However, two of the amusics (Amusics 4 and 6 in Table 1) showed a performance for the scale subtest that was in the normal range (28 and 26, respectively). Furthermore, one amusic (Amusic 1) had very high scores for the rhythm subtest (27). The other four amusics show bad performances for both subtests. Our results thus support Peretz et al.’ assumption of different types of amusia. Based on our results, a dissociation between amelodic and atemporal disorders seems to occur. Studies with larger groups of amusics are necessary to test this hypothesis. If a dissociation is possible, further diagnostics to reliably distinguish the different types of amusia are needed.

V. CONCLUSION

As expected, the amusics in this study performed worse than the control group in detecting pitch differences in non-linguistic stimuli, in line with all previous studies on congenital amusia. The present study further supports the findings by Patel et al. (2008) and Liu et al. (2010) that amusics show an impairment for the perception of intonation (linguistic pitch). Through this, the present study lends support to the hypothesis that amusia is not domain-specific (restricted to music) but a general perceptual impairment.

Continuity and length of the stimuli, the direction of pitch change, and the type of non-linguistic stimuli were all shown to influence the perception of pitch by amusics and non-amusics. While better performances for rising pitch are in line with previous findings and can be attributed to a perceptual-compensation effect, facilitation by discontinuous stimuli and by pulse trains are findings that have to be further explored in future studies.

REFERENCES

- Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia – A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238–251.
- Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.25, retrieved 12 May 2011 from <http://www.praat.org/>
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying “tone deafness”. *Brain*, *127*, 801–810.

- Gosselin, N., Jolicoeur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, 1169, 270–272.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Henry, M. J., McAuley, J. D. (2010). On the prevalence of congenital amusia. *Music Perception*, 27, 413–418.
- Hirst, D., & Di Cristo, A. (1984). French intonation: a parametric approach. *Die Neueren Sprachen*, 83, 554–569.
- Hutchins, S., Gosselin, N., & Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology*, 1, 1–8.
- Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, 15, 356–360.
- Iversen, J. R., & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In K. Miyazaki et al. (Eds.), *Proceedings of the 10th International Conference on Music Perception & Cognition* (pp. 465–468). Adelaide: Causal Productions.
- Jones, J., Lucker, J., Zalewski, C., Brewer, C. & Drayna, D. (2009). Phonological processing in adults with deficits in musical pitch recognition. *Journal of Communication Disorders* 42, 226–234.
- Kalmus, H. & Fry, D. B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics*, 43, 369–383.
- Ladefoged, P. (1997). Linguistic phonetic descriptions. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 589–618). Oxford: Blackwell.
- Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, 133, 1682–1693.
- Lochy, A., Hyde, K., Parisel, S., Van Hyfte, S., & Peretz, I. (2004). Discrimination of speech prosody in congenital amusia. Poster session presented at the meeting of the *Cognitive Neuroscience Society*, San Francisco.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mixdorff, H., & Fujisaki, H. (1995). Production and perception of statement, question and non-terminal intonation in German. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, (vol. 2, pp. 410–413). Stockholm.
- 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.
- Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., & Penhune, V. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, 33, 185–191.
- 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.
- R Development Core Team. (2012). R: A language and environment for statistical computing (Version 2.15.0) [Computer software]. Retrieved from <http://www.r-project.org/>
- Stewart, L. (2008). Fractionating the musical mind: Insights from congenital amusia. *Current Opinion in Neurobiology*, 18, 127–130.
- Thompson, W. F. (2007). Exploring variants of amusia: Tone deafness, rhythm impairment, and intonation insensitivity. In *Proceedings of the Inaugural International Conference on Music Communication Science*, (pp. 159–163). Sydney.
- Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, 71, 259–264.
- Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology*, 6, 15–22.

**Accompanying Files are found in the Proceedings
CD-ROM, in directory /Multimedia/422**