

The Jake Mandell Test as a measure of individual differences in pitch discrimination: validity and reliability properties¹

María-Ángeles Palomar-García², Gustau Olcina-Sempere³, Mireia Hernández⁴, María-Antonia Parcet⁵, Jacob C. Mandell⁶, César Ávila⁷

Recibido: 4 de diciembre de 2019 / Aceptado: 27 de Abril de 2020

Abstract. The Jake Mandell Tone Deaf Test (JMT) is an online measure designed to evaluate pitch discrimination in the normal population. The test has been available online since 2006, but no psychometric data have been published. The present research was designed to provide validity and reliability data for this test. Study 1 was conducted in university students, and the performance on the JMT was compared to the AMUSIA online screening tests. Correlations were high with the subtests of the AMUSIA online screening tests, but especially with the Tone discrimination test. Study 2 was conducted in children, and the JMT was compared to different tests of musical abilities. The JMT showed a good relationship with the Tone Discrimination and Rhythm Imitation subtest, but especially with the global score of musical abilities. In addition, the test showed good one-year test-retest reliability. Finally, external validation of the JMT was obtained by noting that musicians obtained higher scores than non-musicians. Discussion is focused on the possibility of using the JMT as a rapid screening measure of individual differences in pitch discrimination in the normal population.

Keywords: Pitch discrimination; Jake Mandell Tone Deaf Test (JMT); musical abilities; neuroimaging; brain.

Summary. 1. Introduction. 2. Study 1. 2.1 Materials and methods. 2.1.1. Participants. 2.1.2. Materials. 2.2. Results. 2.3. Qualitative analysis. 3. Study 2. 3.1. Materials and methods. 3.1.1. Participants. 3.1.2. Materials. 3.2. Results. 4. Study 3. 4.1. Materials and methods. 4.1.1. Participants. 4.1.2. Materials. 4.2. Results. 5. Discussion. 6. Conclusions. 7. References.

How to cite this article: Palomar-García, M. Á.; Olcina-Sempere, G.; Hernández, M.; Parcet, M. A.; Mandell, J. C. y Ávila, C. (2020). The Jake Mandell Test as a measure of individual differences in pitch discrimination: validity and reliability properties. *Revista Electrónica Complutense de Investigación en Educación Musical*, 17, 143-151.

1. Introduction

Congenital amusia (also known as tone deafness) is the lack of capacity to recognize relative differences in frequency that cannot be explained by hearing loss, brain damage, or cognitive deficits (Ayotte, Peretz & Hyde, 2002; Peretz & Hyde, 2003; Sihvonen, Sarkamo, Rodríguez-Fornells, Ripollés, Munte & Soinila, 2019). This lifelong presence of insensitivity to musical pitch is the main characteristic of congenital amusia, which also requires a preservation of the ability to discriminate different rhythms. Congenital amusia is estimated to be present in about 4% of the population

¹ This research was carried out thanks to a grant from the Spanish Ministry of Economy and Competitiveness (PSI-2016-78805-R) to C. Ávila. Furthermore, this research was made possible by a grant from the postdoctoral postgraduate program (postdoc-UJI to M-Á. Palomar-García) and the Ramón y Cajal Research Program of the Ministry of Science, Innovation and Universities of Spain (RYC-2016- 19477 to M. Hernández).

² University Jaume I
mpalomar@uji.es

³ University Jaume I
golcina@uji.es

⁴ University of Barcelona
mireiahp@gmail.com

⁵ University Jaume I
parcet@psb.uji.es

⁶ Brigham and Women's Hospital
jmandell@partners.org

⁷ University Jaume I
avila@psb.uji.es

(Peretz & Vuvan, 2017). Neuroimaging studies have associated this disorder with modifications in gray matter volume in the superior temporal gyrus and/or inferior frontal gyrus, interpreted as an alteration in the brain systems related to auditory processing and memory (Albouy et al., 2013; Sihvonen, Ripollés, Leo, Rodríguez-Fornells, Soinila & Särkämö, 2016). Furthermore, congenital amusia is related to abnormal connectivity in fronto-temporal network, configuring as a disconnection syndrome.

The evaluation of tone deafness and the diagnosis of congenital amusia have mostly been carried out with the Montreal Battery for Evaluation of Amusia (MBEA; 2). The development of this battery was based on the cognitive-neuropsychological model developed by Perez and Coltheart (Peretz & Hyde, 2003). This model predicts the existence of separate auditory modules for language and music processing. The module for music processing is formed by two independent and parallel systems, one for melodic processing (more related to pitch discrimination) and the other for temporal organization (more associated with the rhythm processing). These two systems were hypothesized based on data obtained in patients with lesions in the superior temporal gyrus who showed impaired pitch discrimination in the presence of a preserved capacity for variations in rhythm (Peretz, Kolinsky, Tramo, Labrecque, Hublet, Demeurisse & Belleville, 1994; Piccirilli, Sciarma & Luzzi, 2000).

The MBEA is composed of six different tests, three dedicated to different aspects of pitch discrimination, one for rhythm perception, one to discriminate whether the melody was a march or a waltz, and the last one for melody recognition. The diagnosis of amusia requires a low score on the three pitch discrimination tasks, whereas rhythm perception abilities are preserved (Vuvan, Paquette, Goulet, Royal, Felezeu & Peretz, 2018). The MBEA has demonstrated its capacity to detect congenital amusia in studies with populations of individuals with brain injuries and with healthy populations (Ayotte et al., 2002; Peretz et al., 2002). The psychometric properties of the MBEA were acceptable. Test-retest reliability was 0,75 at 4 months, and concurrent validity was good when correlated with the scores on Gordon's Musical Aptitude Profile tests [$r = 0.53$; (Peretz et al., 2002)]. However, the diverse tests of the MBEA were easy for most of the population, as the distribution of scores was asymmetric, with a tendency toward high scores (Peretz, Champod & Hyde, 2003).

Another test, focused on overall pitch perception ability, was developed by Jake Mandell in 2006 and has been available online for more than 10 years: the Jake Mandell Tone Deaf Test (JMT). This test was initially developed to evaluate tone-deafness (congenital amusia), but at the same time it was designed to be challenging, even for subjects with musical training. This difficulty level was achieved by using complex and varied longer melodies with multiple musical timbres where one single note may change. Thus, the test was designed to reduce any potential bias due to specific instrument training and, prevent the clustering of high scores by trained individuals. The informal data obtained online from 61.036 subjects is presented on the webpage and show that this instrument may be useful to evaluate the capacity for pitch perception. The distribution of scores followed a normal curve, with mean percentage of 73.8% ($SD=9.99$), and very few respondents reached scores over 90%, which makes the JMT an interesting test to measure musical abilities. The presence of complex and longer melodies increases the requirements for pitch memory (Gosselin, Jolicoeur & Peretz, 2009; Tillmann, Schulze & Foxton, 2009). This process has been altered in congenital amusia, and it has also been related to the auditory cortex and opercular frontal areas (Albouy et al., 2013). This previous literature showed that both pitch perception and pitch memory share a common brain network in bilateral frontotemporal areas.

The contribution of the present study is to provide reliability and concurrent validity data for the Jake Mandell test, in order to demonstrate its feasibility as a measure of individual differences in pitch discrimination and pitch memory. The online data obtained for the test suggest that the use of complex stimuli makes the test more difficult than other pitch tests, but no study has investigated this possibility. Our specific objectives are: 1) to study the internal consistency and stability of JMT; 2) to compare the JMT to other tests measuring musical abilities; and 3) to show the performance on the JMT in participants with musical training. The development of this instrument will allow the existence of a fast and psychometrically reliable measure for the evaluation of tonal discrimination capacities that could be useful in the educational field and in the evaluation of musical abilities in the clinical field.

Three different studies were carried out. The first study was conducted in university students and designed to investigate concurrent validity with AMUSIA online screening tests (online version; Peretz & Vuvan, 2017; Vuvan et al., 2018). This screening tool is composed of three tests: the Scale test, the off-beat test, and the out-of-key-test; which measure pitch discrimination, time-based deafness, and off-key notes, respectively. Compared to AMUSIA online screening tests, we expected to find that the JMT showed a stronger correlation with AMUSIA online screening tests (especially with the Scale test, which is very similar), a more symmetric distribution of scores, and lower scores due to its greater difficulty. Study 2 was conducted in children, and we expected to obtain a similar distribution as in adults and specific correlations with tests of tone discrimination. Finally, in the third study we compared performance on the JMT in musicians and non-musicians. We expected to find a symmetric distribution of scores in both groups, but higher scores on the test for the expert group.

2. Study 1

2.1. Materials and methods

2.1.1. Participants

Our sample was composed of 61 university students (34 females). The age of the participants ranged from 18 to 38 years (mean 22.41; SD=3.64). None of the participants had ever played a musical instrument, and they had received no musical training beyond normal school education. Written informed consent was obtained from all the participants, following a protocol approved by the Universitat Jaume I, and they received monetary compensation.

2.1.2. Materials

1. Jake Mandell Tone Deaf Test (JMT)

This computer test consists of 36 trials based on complex musical phrases that use different sonorities, such as organ, piano, percussion or string instruments. These phrases were also heterogeneous in different features, such as duration, number of tones, number of short and long sounds, changes in intensity, and the use of synthesized or natural sounds. The trials are comprised of paired brief musical phrases performed in a variety of timbres and musical styles, with each matched pair sharing the same melodic contour, rhythm, and timbre. Half of the pairs (18/36) differ in the pitch of a single note, with 9/18 of the different notes falling outside of the scale of the melody and 9/18 confined to the scale. The pitch difference of the single modified note from the initial and repeated phrase may vary by up to 11 semitones in pitch; variations greater than one octave were not utilized. These changes always occurred in one of the last 10 tones of the melody.

On each trial, the subject hears two short successive melodies and indicates whether they are equal (green button “same”) or different (red button “different”); pairs of melodies can be the same or have differences in one or more pitches. After receiving the instructions, the subject is given the opportunity to adjust the volume to a comfortable level, and four test trials are presented. Then 36 paired trials are presented to all subjects in the same order, without arranging the items in the order of increasing difficulty; half of the pairs of trials are different melodies, and the other half are the same. This test is available at <http://jakemandell.com/tonedeaf>. We used the percentage of correct answers as scores in our analysis. Based on these scores, possible pitch discrimination deficit was considered for scores below 55%.

2. AMUSIA online screening tests (Peretz & Vuvan, 2017; Vuvan et al., 2018)

This test is performed with a computer via the internet (<http://www.brams.org/amusia-public/>), and it includes different subtests: the Scale subtest extracted from the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2002) and two subtests based on the MBEA scales, the Off-beat subtest and Out-of-key subtest (Peretz et al., 2008).

On the Scale subtest, participants performed 31 trials. Each trial presents two short successive melodies, and subjects must indicate whether they are the same or different melodies by clicking on the “different” or “same” button. The melodies that are different differ in one pitch. The off-beat subtest contains 24 trials. Each trial consists of listening to a short melody. Subjects have to identify whether or not an unusual delay appears in the melody: if it does, they click on “incongruous”; if it does not, they click on “correct”. This delay is a silence of 5/7 of the beat duration (i.e., 357 ms) prior to the critical tone. The out-of-key subtest contains 24 trials. Each trial consists of listening to a short melody and identifying whether it contains a note that is off the scale of the melody by clicking on “incongruous” or “correct”. All the participants perform the subtests in the same order (Scale, Off-beat, and Out-of-key). Before starting, each subtest was verbally explained to the participants, who performed two to four practice trials for each subtest and adjusted the volume of the headphones to clearly hear the melodies. On each subtest, half of the trials belong to one condition and the other half to the other condition and trials are presented in a random order.

We used percentages of correct answers as scores on each test in our analysis. Based on these scores, the possible pitch discriminant deficit (i.e. amusia) was considered for scores below 70% on Scale and Out-of-key, and a score over above 70% on the Off-beat test (Vuvan et al., 2018).

2.2. Results

Mean percentage of correct responses, standard deviations, and range for each test are reported in Table 1. Paired t-tests among the four tests were conducted to study differences between them. Results showed that the Jake Mandell Test was more difficult than the three AMUSIA online screening tests. There were no gender differences on any of the tests ($P > 0.10$). Pearson’s correlations with age were all non-significant ($P > 0.10$).

Table 1. Mean percentage of correct responses, standard deviations, and range for the JMT and the AMUSIA online screening tests, as well as paired t-tests between them

Test	Mean (SD)	Range	Paired t-tests
1. JMT	65.43 (8.70)	47-86.1	1 < 2; t (59) =8.66, p < .001 1 < 3; t (59) =8.25, p < .001 1 < 4; t (59) =3.16, p =.002
2. Scale	76.93 (10.53)	57-100	2 > 4; t (59) =3.01, p = .003
3. Off-beat	76.02 (8.85)	54-92	3 > 4; t (59) =2.32, p = .02
4. Out-of-key	71.58 (15.05)	46-96	

Internal consistency of the JMT was measure using McDonald's ω_H reliability because was able to outperform Cronbach's alpha (Zinbarg, Revelle, Yovel & Li, 2005). McDonald's ω_H describes the variance of the total score that can be attributed to the general factor (Hermsen, Leone, Smalbrugge, Knol, van der Horst & Dekker, 2013). The McDonald's ω_H obtained for the present sample was 0.91, which should be interpreted as support for the unidimensionality of the measure.

We also ran Pearson's correlations between the different tests (Table 2). As expected, the JMT correlated significantly with the Scale test on the MBEA as well as with the other two MBEA tests.

Table 2. Pearson's correlations between scores obtained on the JMT and the AMUSIA online screening tests

	Scale	Off-beat	Out-of-key
1. Jake Mandell Test	.44**	.33**	.38**
2. Scale		.28*	.51**
3. Off-beat			.35*
4. Out-of-key			

**P < 0.01; *P < 0.05

2.3. Qualitative Analyses

According to the JMT criteria, amusia may be suspected when accuracy on the test is lower than 55%, and 11.4% of our sample met this criterion. Similarly, the AMUSIA online screening tests required scores of less than 70% on the Scale and Out-of-key tests and scores greater than 70% on the Off-beat test, and 11.5% of our sample met this criterion. The agreement between the two measures reached 85.2% (Cohen's Kappa= 2.48, p =.013), with 3 cases out of 7 diagnosed as amusic by both measures.

3. Study 2

3.1. Materials and methods

3.1.1. Participants

Our sample was composed of 33 children from different private and public schools of Castellon with no formal previous musical instruction who were interested in accessing formal music studies in the conservatory. The group was made up of 15 boys and 18 girls. The age of the participants ranged from 6 to 12 years (M=8.47; SD=1.50). Participants had never played a musical instrument and they had received no musical training beyond normal school education. Parents gave their written consent for their children to participate in this study, following a protocol approved by the University Jaume I, and they received economic compensation for their participation.

3.1.2. Materials

Participants completed the JMT plus other tests developed to specifically assess five different musical aptitudes: Tone Discrimination, Rhythmic Memory, Tonal Memory, Rhythmic Imitation, and Melodic Imitation. These aptitudes

were inspired in Bentley's Musical Aptitude Test (Young, 1973), but we decided to use shorter forms of Bentley's tests for two reasons: 1) these shorter versions are frequently used to select children for entry into the conservatory in Spain; and 2) the participants were involved in a more global neuroimaging study, and we did not enough time to use longer tests. The items on each test were recorded, interpreted, administered, and evaluated by a professional musician. More detailed information can be requested from the authors.

1. The Tone Discrimination test requires participants to listen to 10 items with two notes each played with a flute. On 5 items, the second sound rises, whereas on the other 5 items, the second sound falls. On this task, the child should recognize whether the second note goes up or down relative to the first one. The score for this test was the sum of all correct responses (maximum score = 10).
2. The Rhythmic Memory test consists of 10 different musical rhythms with four beats. Participants listened to two consecutive repetitions each of these rhythms. On 50% of the trials, the beats in the first and second rhythm repetitions were the same, whereas a note was modified on the other 50%. Thus, the participants had to indicate whether or not the beat of the second rhythm was the same or different from the first one. If the participant noticed any change, he/she was asked to identify what beat had changed. The score for this test was the sum of all correct responses (maximum score = 10).
3. The Tonal Memory test consists of 10 different musical melodies with five beats each. Participants listened to two consecutive repetitions of each of these melodies. The investigators altered one note in 5 of the 10 original musical melodies in the second repetition. Thus, the participants had to identify whether each melody was the same or different from the first one. Then, if the participant noticed any change, he/she was asked to identify what beat had changed. The score for this test was the sum of all correct responses (maximum score = 10).
4. The Rhythmic Imitation test consists of 5 different hand-clapping rhythms. First, the examiner showed participants how to perform a hand-clapping rhythm without giving them a visual guide. Then, participants listened to the 5 hand-clapping rhythms (with no visual guide) and had to imitate them. The score for this test was the sum of all correct responses (maximum score = 10).
5. The Melodic Imitation test is composed of 5 different musical phrases. The examiner sings each of the 5 musical phrases with the syllable "la", and the children have to sing the melodies back. The score for this test was the sum of all correct responses (maximum score = 10).
6. Global score. We calculated a global score for music abilities by adding standard z-scores (Mean=0; SD=1) obtained from the total scores on each of the previous tests.

3.2. Results

Means and standard deviations for all tests are reported on Table 3. The JMT scores were lower in children than those obtained by adults in Study 1 (difference = 4.07), but the difference only approached significance ($t=1.91$; $P = 0.058$). The JMT scores followed a normal distribution (K-S test $Z=0.73$, $P > 0.10$). Partial correlations (controlling for age) between different measures are reported on Table 3. As expected, the JMT correlated with the other variables, especially with the Tone Discrimination and the Rhythm Discrimination tests. Importantly, the partial correlation of the JMT with the Global Score on the musical tests was strongly significant.

Table 3. Partial correlations between scores obtained on the JMT and the different musical tests, as well as mean correct responses, standard deviations, range and Pearson's correlations with age

	Jake Mandell Test	Tone Discrimination Test	Rhythmic Memory Test	Tonal Memory Test	Rhythmic Imitation Test	Melodic Imitation Test
JMT	1					
Tone Discrimination Test	.38*	1				
Rhythmic Memory Test	.33	.45*	1			
Tonal Memory Test	.24	.27	.40*	1		
Rhythmic Imitation Test	.40*	.21	.44*	.51**	1	
Melodic Imitation Test	.30	.18	.29	.20	.30	1
Global Score	.56**					

Mean (SD)	61.36 (9.3)	8.32 (1.70)	8.59 (1.40)	8.71 (1.19)	6.35 (3.25)	2.94 (2.44)
Range	38-86	3-10	6-10	6-10	0-10	0-10
Pearson's r with age	.28	.42*	.22	.26	.18	.28

*P <0.05; **P <0.01

One-year test-retest reliability was calculated with the whole sample. Results showed that scores improved after one year (M= 66.25; SD=9.73), and the intraclass correlation (ICC) between the two measures was high [ICC = 0.59 (95% CI 0.172-0.80); see Fig 1].

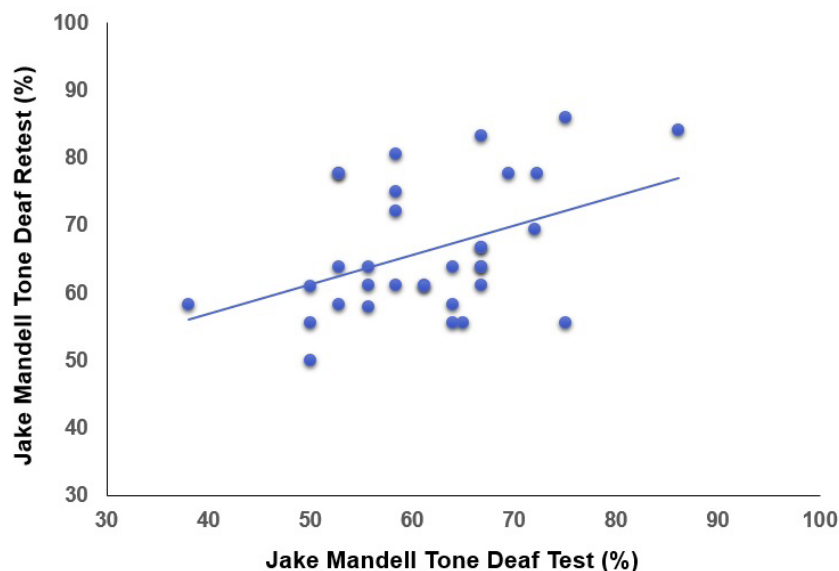


Fig. 1. The scatter plot shows the intraclass correlation (ICC) between the scores obtained on the Jake Mandell Test with scores obtained on the Retest

Moreover, the McDonald's ω_H obtained for the present sample was 0.81, which should be interpreted as support for the unidimensionality of the measure.

4. Study 3

4.1. Materials and methods

4.1.1. Participants

A total of 61 voluntary subjects that responded to advertisement participated in the study: 33 musicians and 28 non-musicians. They responded to add Musicians had completed formal music studies (conservatory, private schools) for at least 9 years, and they were active musicians (11 females; mean age=20.09 yrs; SD=2.01; range: 18-26 yrs). Non-musicians had only received obligatory musical instruction at school (12 females; mean age=20.68 yrs; SD=2.21; range: 18-27 yrs). The two groups did not differ in age or gender distribution. Written informed consent was obtained from all the participants, following a protocol approved by the Universitat Jaume I, and they received monetary compensation.

4.1.2. Materials

Participants completed the JMT (see Study 1 for methodological details).

4.2. Results

Mean percentage of correct responses and standard deviations of both groups are reported on Table 4. An ANOVA conducted using the JMT as dependent variable, Gender and Group as between-subjects factors, and age as a covariate only yielded a significant effect for group [$F(1, 36) = 12.68, P = .002, \eta^2 = .40$]. As expected, musicians obtained better scores than non-musicians.

Table 4. Mean correct responses and standard deviations on the JMT in the musician and the non-musician groups

Group	N	Mean	SD	Typical error of the mean
Non-musicians	28	67.85	10.43	1.97
Musicians	33	76.84	9.43	1.64

5. Discussion and conclusions

In the present research, psychometric data from the JMT in adults and children have been presented to validate this instrument as a measure of individual differences in pitch discrimination. Studies 1 and 2 showed good internal consistency of the scale and good convergent validity when JMT scores were correlated with other variables that tapped similar musical abilities. Importantly, mean scores on the JMT were lower than those obtained for the AMUSIA online screening tests, showing a more symmetric and normal distribution. Study 2 also showed that the JMT had good test-retest reliability and was related to global musical abilities. Finally, Study 3 gave more content validity to the test by showing that musicians had better performance than non-musicians, but without finding a ceiling effect. Overall, the present data are consistent with the idea that the JMT is a good psychometric, rapid instrument for measuring individual differences in pitch perception and memory.

Study 1 was conducted in non-musicians adults and found a mean score on the JMT of 65.4%, 10% less than the mean score found in the online sample. The scores obtained were distributed symmetrically across the participants in a range from 47 to 86%. In this group of participants, 11.5% were considered amusic according to the criteria proposed by the test (scores below 55%). It is worth nothing that, the JMT was also sensitive to identifying individuals with good pitch discrimination abilities because there was no ceiling effect: none of the tested participants reached more than 87% accuracy.

Concurrent validity of the JMT was studied using three different tests included in the AMUSIA screening tests. As expected, the JMT mean was lower than those for the three AMUSIA online screening tests. Correlation analyses revealed a strong association between the JMT and the Scale subtest of the AMUSIA online screening test. The Scale test has a similar format as the JMT, but it is easier and, thus, leads to ceiling effects. Some studies only used this Scale test to estimate possible deficits in music perception (Sihvonen et al., 2016; Peretz & Vuvan, 2017). Moreover, the JMT showed significant but lower correlations with the other two AMUSIA tests: the Out-of-key and Off-Beat tests. These results were consistent with previous reports showing significant correlations between all the MBEA tests (Pfeifer & Hamann, 2015), even the requirement for amusics to have a preserved perception of rhythm. In fact, when factorial analysis of the different MBEA tests was performed, a one-factor model was obtained, indicating that the different musical abilities were all correlated. Thus, the JMT may also be considered a measure of general musical abilities.

The JMT was administered to 6-12 years old children in Study 2. The mean score was approximately 4 points lower than in adults, with a range between 38 and 86%. Differences between children and adults and correlations with age within the sample of children were non-significant. This suggests that the development of pitch discrimination in non-musicians as measured with the JMT did not vary with age and may be established during childhood at 7 years old. Indeed, previous research showed that these abilities are mostly developed until the age of 7, with no differences found between childhood and young adolescents (Thompson et al., 1999). Our results were consistent with this previous report. The JMT also showed a good test-retest reliability index at one year, suggesting that test performance is stable, even though the second administration yielded an improvement in performance (of less than 5%).

We also studied concurrent validity in children by correlating the JMT scores with the results of five different constructed measures. As expected, the correlation was positive with all the measures, but it was only significant with the Tone Discrimination test and the Rhythmic Imitation test. As in Study 1, the Tone Discrimination test was a similar and easier version of the JMT that required participants to indicate whether two consecutive melodies were the same or different. The Rhythmic Imitation test required participants to clap an auditory melody presented, and so good pitch memory was also required. The correlations with the other three subtests were positive, but they only approached significance, indicating that some musical perception abilities were required. Importantly, the strongest correlation was observed with a composite factor of all tests, indicating that the JMT was associated with all music abilities. Thus, the JMT could reflect a global measure to evaluate all musical abilities related to pitch perception and memory. In this regard, the JMT has shown a good association with increased gray matter volume in fronto-temporal areas (superior temporal gyrus and inferior frontal gyrus) in both children and adults (Palomar-García et al., 2020).

The third study was designed to evaluate predictive validity by studying JMT performance in a group of musicians. Musicians typically show an enhanced pitch discrimination ability compared to non-musicians, consistent with the finding that musicians are more sensitive to some acoustic features that are critical for music processing (Spiegel & Watson, 1984; Micheyl et al., 2006). Consistently, expertise in music processing and performance was associated with higher scores on the JMT, compared to non-musicians. The scores were almost 10% higher than those of non-musicians. However, the scores of musicians also followed a normal distribution, indicating the lack of a ceiling

effect. Importantly, scores on the JMT did not correlate with earlier musical training or with years of musical training, suggesting a possible independence of training.

Overall, the present studies suggest that pitch discrimination abilities are established at 6-7 years of age and barely improve in the period from childhood to adulthood without specific training. The lack of correlation with age and the lack of significant differences between children and adults suggest the absence of maturational effects during development. Thus, we cannot conclude that the better performance of musicians on the JMT is due to training because a self-selection process may cause individuals with good pitch discrimination abilities to become musicians.

The current instrument has several potential uses. First, it should help to rapidly measure pitch discrimination abilities and general musical abilities in healthy adults, and to establish individual differences among musicians. Second, JMT may be useful as a tool to investigate music perception development in children. Finally, JMT could be useful in the educational field as a measure of the potential of musical ability, as well as in the clinical field as a measure of musical processing deficits associated with brain injuries.

6. References

- Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A. & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*, 136(5), 1639-1661.
- Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125(2), 238-251.
- Bianchi, F., Santurette, S., Wendt, D. & Dau, T. (2016). Pitch discrimination in musicians and non-musicians: effects of harmonic resolvability and processing effort. *Journal of the Association for Research in Otolaryngology*, 17(1), 69-79.
- Gosselin, N., Jolicoeur, P. & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, 1169(1), 270-272.
- Hermesen, L. A., Leone, S. S., Smalbrugge, M., Knol, D. L., van der Horst, H. E. & Dekker, J. (2013). Exploring the aggregation of four functional measures in a population of older adults with joint pain and comorbidity. *BMC geriatrics*, 13(1), 119.
- Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C. & Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *Journal of Neuroscience*, 27(47), 13028-13032.
- Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P. & Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain*, 129(10), 2562-2570.
- Kishon-Rabin, L., Amir, O., Vexler, Y. & Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*, 12(2), 125-144.
- Mandell, J., Schulze, K. & Schlaug, G. (2007). Congenital amusia: an auditory-motor feedback disorder? *Restorative Neurology and Neuroscience*, 25(3-4), 323-334.
- Micheyl, C., Delhommeau, K., Perrot, X. & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1-2), 36-47.
- Palomar-García, M. Á., Hernández, M., Olcina-Sempere, G., Adrián-Ventura, J., Costumero, V., Miró-Padilla, A., Villar-Rodríguez, E., & Ávila C. (under review). Auditory and frontal anatomic correlates of pitch discrimination in musicians, non-musicians, and children without musical training. *Brain Structure and Function*.
- Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B. & Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron*, 33(2), 185-191.
- Peretz, I., Champod, A. S. & Hyde, K. (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, I., Gosselin, N., Tillmann, B., Gagnon, B., Trimmer, C. G., Paquette S. & Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception: An Interdisciplinary Journal*, 25(4), 331-343.
- Peretz, I. & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, 7(8), 362-367.
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G. & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117(6), 1283-1301.
- Peretz, I. & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics*, 25(5), 625-630.
- Piccirilli, M., Sciarra, T. & Luzzi, S. (2000). Modularity of music: evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery & Psychiatry*, 69(4), 541-545.
- Pfeifer, J. & Hamann, S. (2015). Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia. *Frontiers in Human Neuroscience*, 9, 161.
- Sihvonen, A. J., Ripollés, P., Leo, V., Rodríguez-Fornells, A., Soinila, S. & Särkämö, T. (2016). Neural basis of acquired amusia and its recovery after stroke. *Journal of Neuroscience*, 36(34), 8872-8881.
- Sihvonen A. J., Sarkamo T., Rodríguez-Fornells A., Ripollés P., Munte T., F. y Soinila S. (2019). Neural architectures of music - Insights from acquired amusia. *Neuroscience & Biobehavioral Review*, 107, 104-114.
- Spiegel, M. F. & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, 76(6), 1690-1695.

- 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(3), 259-264.
- Thompson, N. C., Cranford, J. L. & Hoyer, E. (1999). Brief-tone frequency discrimination by children. *Journal of Speech, Language, and Hearing Research*, 42(5), 1061-1068.
- Vuvan, D. T., Paquette, S., Goulet, G. M., Royal, I., Felezeu, M. & Peretz, I. (2018). The montreal protocol for identification of amusia. *Behavior Research Methods*, 50(2), 662-672.
- Young, W. T (1973). The Bentley "Measures of Musical Abilities": A Congruent Validity Report. *Journal of Research in Music Education*, 21(1): 74-79.
- Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123-133.