Auditory Perception & Cognition

ISSN: 2574-2442 (Print) 2574-2450 (Online) Journal homepage: https://www.tandfonline.com/loi/rpac20

Testing Convergence between Singing and Music Perception Accuracy Using Two Standardized Measures

Peter Q. Pfordresher & Nicholas P. Nolan

To cite this article: Peter Q. Pfordresher & Nicholas P. Nolan (2019) Testing Convergence between Singing and Music Perception Accuracy Using Two Standardized Measures, Auditory Perception & Cognition, 2:1-2, 67-81, DOI: 10.1080/25742442.2019.1663716

To link to this article: https://doi.org/10.1080/25742442.2019.1663716

Published online: 09 Sep 2019.

AUDITORY PERCEPTION & COGNITION
2019, VOL. 2, NOS. 1–2, 67–81
https://doi.org/10.1080/25742442.2019.1663716

Routledge
Taylor & Francis Group

Testing Convergence between Singing and Music Perception Accuracy Using Two Standardized Measures

Peter Q. Pfordresher and Nicholas P. Nolan

Department of Psychology, University at Buffalo, State University of New York, Buffalo, USA

ABSTRACT

Research on individual differences in musical abilities, and music-related deficits, has increased dramatically in the past 20 years. Although most studies to date concern music perception, in particular the deficit referred to as *congenital amusia*, a growing area of research has addressed individual differences in singing accuracy and *poor-pitch singing*. How closely associated are music perception and singing abilities? Several studies to date have reported dissociations between these abilities. However, these studies have tended to use small samples and have not compared leading standardized measures to each other. In the present study, we measured perception and singing abilities in a larger sample (N = 86) on standardized measures of singing accuracy (the Seattle Singing Accuracy Protocol) and music perception (the Online Test of Amuisa). Results revealed stronger associations between these higher-level musical abilities than either measure had with simple pitch discrimination. Analyses in which participants were classified as typical or deficient based on existing norms further suggested that deficits in music perception predict poor-pitch singing deficits more so than the reverse. Taken together, these results suggest that music perception and production may rely on shared higher-order representations of music that play a more important role than basic perception or motor control.

ARTICLE HISTORY
Received 18 February 2019
Accepted 8 August 2019

KEYWORDS
Singing accuracy; perception and action; music cognition; musical deficits

Singing is the most widespread form of music performance. Culturally, music and singing facilitate group cohesion in social gatherings through the usage of national anthems, songs like *Happy Birthday*, and chants in religious services (Clift et al., 2010; Stewart & Lonsdale, 2016; Welch, Himonides, Saunders, Papageorgi, & Sarazin, 2014). Vocalizing during infancy plays an important role in the development of effective communication (Welch, 2005). Nevertheless, certain individuals exhibit deficient singing accuracy. These individuals are often referred to as *poor-pitch singers* (Welch, 1979), and may suffer from an underlying *Vocal Pitch Imitation Deficit* (VPID, Pfordresher & Larrouy-Maestri, 2015) that hinders vocal pitch imitation of speech as well as music (Mantell & Pfordresher, 2013; Wisniewski, Mantell, & Pfordresher, 2013).

One plausible cause for poor-pitch singing originates in the colloquial term "tone deaf." Although this term is usually associated with singing production, rather than

CONTACT Peter Q. Pfordresher pqp@buffalo.edu University at Buffalo, State University of New York, Buffalo, USA

perception (Cuddy, Balkwill, Peretz, & Holden, 2005), it hints at a possible perceptual basis (the inability to perceive tones musically) for poor singing. Such an account would be consistent with a leading model of singing accuracy, the *Vocal Sensorimotor Loop* (VSL, Berkowska & Dalla Bella, 2009), illustrated in Figure 1. According to the VSL, a deficit in perception should necessarily lead to problems in auditory-motor mapping, thus causing motor control of pitch to be inaccurate. In addition, this model predicts that problems of singing accuracy may exist without concurrent deficits in perception. An individual could, for instance, experience a deficiency specific to auditory-motor mapping or motor planning without a deficit in perception. Such an individual should succeed on music perception tests while failing to sing accurately.

In short, the VSL predicts that poor-pitch singing should emerge as a necessary consequence of *congenital amusia* (Peretz et al., 2002). At present congenital amusia is defined as "lack of awareness of acquired musical pitch knowledge" (Peretz, 2016, p. 857) that cannot be explained by hearing loss, brain damage, intellectual deficits or lack of musical exposure. In the present study, we addressed whether individuals who may be classified as exhibiting congenital amusia are likely to exhibit an additional pitch production deficit, as predicted by the VSL. It is important to note that this model predicts associations based on categorical distinctions as well as how people score on a continuum, with individuals scoring relatively poorly on music perception tests also performing poorly on singing accuracy, even while not exhibiting a classifiable "deficit" on either measure.

Evidence to date is mixed with respect to support for the VSL model and generally favors dissociations between pitch perception and singing. Although an earlier study of congenital amusia found that listeners rated sung performances by this group as poor compared to healthy controls (Ayotte, Peretz, & Hyde, 2002), other studies of congenital amusia have identified cases in which singing is spared (Dalla Bella, Giguère, & Peretz, 2009), as well as evidence that congenital amusic production correlates with

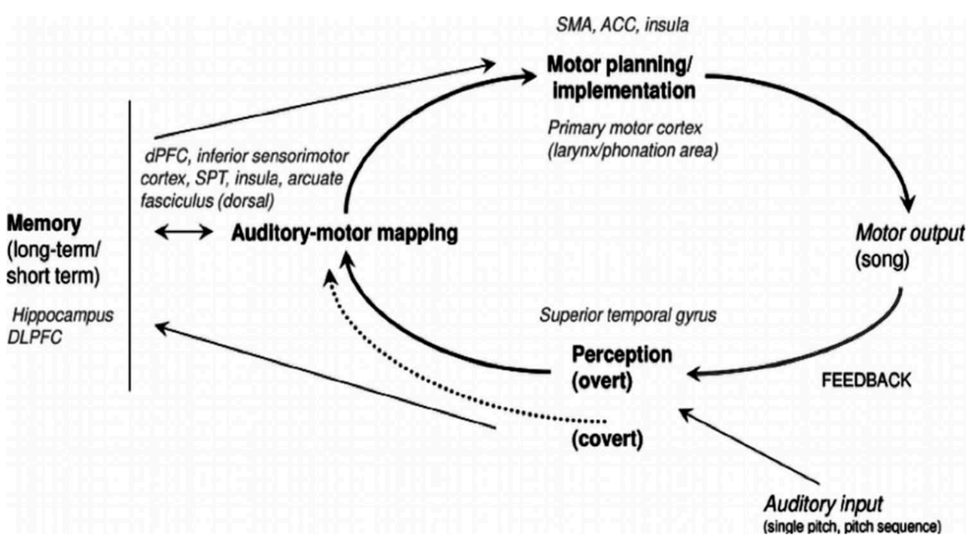

Figure 1. The sensorimotor vocal loop (Berkowska & Dalla Bella, 2009).

pitch height whereas a model like the VSL would predict a breakdown in this association for amusics (Hutchins, Zarate, Zatorre, & Peretz, 2010). Other results suggest a more exaggerated dissociation. Loui and colleagues found that a group of congenital amusics were able to reproduce the direction of a pitch change vocally but were at chance when verbally labeling the direction of that change as upward or downward (Loui, Guenther, Mathys, & Schlaug, 2008; but see Williamson, Liu, Peryer, Grierson, & Stewart, 2012). Studies focusing on singing accuracy likewise have reported low or nonsignificant correlations between singing accuracy and pitch discrimination accuracy (Greenspon, Pfordresher, & Halpern, 2017; Pfordresher & Brown, 2007), no association between singing accuracy and detection of deviations in melody tones (Dalla Bella, Giguère, & Peretz, 2007), and dissociations between pitch matching via singing versus other tasks which may involve more perceptually-based adjustments (Demorest, 2001; Hutchins, Larrouy-Maestri, & Peretz, 2014; Hutchins & Peretz, 2012).

A prevailing weakness in the studies cited above is that none of them involve measurements of singing versus music perception that are matched with respect to the complexity of the task and/or the granularity of analyses. For instance, Pfordresher and Brown (2007) compared accuracy of singing melodies to simple pitch discrimination. Whereas tonal knowledge likely contributed to the singing task it may not have contributed to the perception task. Ayotte and colleagues (2002) compared perceptual tasks involving tonal relationships among pitches in a melody to accuracy of song singing; however, analyses of singing were based on subjective ratings rather than the acoustic signal. Moreover, recent developments allow researchers to compare performance on standardized tasks designed to measure perception and singing optimally.

The present research capitalizes on the presence of standardized measures of both music perception and singing accuracy. For some time, a standardized measure of music perception has been used to diagnose cases of congenital amusia. The Montreal Battery for the Evaluation of Amusia (MBEA, Peretz, Champod, & Hyde, 2003) includes several subtests designed to test the perception of musical pitch, rhythm, meter, and memory. The most informative tests for diagnosing amusia are those that measure the perception of musical pitch. Each of these tests comprises trials in which a novel melody is repeated twice, with an alteration to one of the pitches in the second presentation on half the trials. Most of these alterations are easy to detect for non-amusic listeners, whereas amusics typically only detect fewer than 70% of changes. The present research incorporated a shorter Online Test of Amusia (OTA, Peretz et al., 2008; Peretz & Vuvan, 2017) that can be used for a tentative initial evaluation of amusia (Vuvan et al., 2018).

More recently a standardized measure of singing accuracy was developed that can be used to diagnose poor-pitch singing, the Seattle Singing Accuracy Protocol (SSAP, Demorest & Pfordresher, 2015; Demorest et al., 2015; Pfordresher & Demorest, in press), which was designed to contain a minimal set of tasks to measure pitch accuracy in singing. The most critical tasks in this protocol involve the imitative production of single pitches and 4-note tonal melodies. The pitches used are matched to an individual's pitch range through an automated range-finding procedure.

The closest approximation to the present study was reported in an unpublished doctoral dissertation by Karen Wise (Wise, 2009). She compared the accuracy of singing novel melodies to performance on the MBEA and reported significant correlations with a sample size of 50, including 14 participants who were classified as congenital amusics (though relationships were still significant when these participants were removed from the sample). Wise also found strong associations between the accuracy of imitating pitch intervals and psychophysical pitch discrimination (i.e., correctly labeling the direction of a pitch change). The present study builds on this unpublished work to compare performance on the OTA to performance on the SSAP, which includes tasks similar to Wise but in a standardized format, with a larger sample. In addition, we use regression analysis to determine the relative contribution of pitch discrimination versus the OTA in predicting singing accuracy, and use norm-based cutoffs from the SSAP and OTA to draw conclusions about the degree to which congenital amusia may be associated with VPID. In this way the present research investigates associations between perception and singing accuracy based on both continuous measures and discrete categorization of musical deficits.

Methods

Participants

One-hundred college-age participants (45 female, *M* age 19.7) volunteered to complete this study in exchange for course credit as part of the participant pool associated with the Introduction to Psychology course at the University at Buffalo, The State University of New York. The only criteria for recruitment were normal hearing and vocal abilities which were assessed via self-report. Following removal of participants with problematic pitch discrimination scores (see *Results*), the final sample included 86 participants (39 female *M* age 19.6). The mean years of musical training was 2.5 (range 0–20); only 26 participants reported one or more years of training (*M* within this group = 7.9 years). Most musicians in the sample were instrumentalists; only three participants reported formal training in singing (2 participants reported 3 years, and a third reported 8 years).

Apparatus

Participants were seated in front of an Acer S200HQL 20-inch LED computer monitor connected to a 3.6 GHz PC running Windows 10. Vocal productions were recorded using a Shure WH30 headset microphone connected through a Lexicon Omega I/O box. Auditory stimuli were presented through Sennheiser HD 280 Pro headphones placed over the headset microphone.

Materials

Seattle Singing Accuracy Protocol (SSAP)

The SSAP is an internet-based program that includes automated scoring of pitch accuracy (Demorest & Pfordresher, 2015; Demorest et al., 2015). The first section of the protocol comprises vocal warm-up exercises that are used to determine an optimal key for the participant's voice. This is followed by a series of vocal imitation trials similar to those used by Wise (2009). Next participants sing a familiar song (also used during warm-up) with lyrics and then on the syllable "doo". Participants then complete an adaptive pitch discrimination task designed to identify their discrimination threshold. Finally, participants complete self-report items concerning demographics, their musical background and their musical self-image. The two most critical sections of the SSAP for the present research are the vocal imitation trials and pitch discrimination task. We also analyzed participants reported years of formal musical training on instrument or voice.

Vocal imitation trials included three types of stimuli clustered into subsections: imitation of single notes based on a vocal timbre (10 trials), imitation of single notes based on a piano timbre (10 trials), and imitation of four-note melodies based on a vocal timbre (6 trials). All of these sections were presented using a call-and-response procedure in which participants first listen to a recording of the target stimulus and then attempt to reproduce it by singing on the syllable "doo". Pitches spanned a musical perfect fifth (seven semitones) within the participant's optimal key (established during warm-up). Inter-onset intervals for melodies were 1 second, and the duration of each sung pitch was approximately 800 ms.

The pitch discrimination procedure was based on Loui et al. (2008). On each trial, two pitches were presented in succession and participants were asked to indicate whether the second pitch they heard is higher or lower than the first, with pitch direction being selected at random. The pitch discrimination task used an adaptive three up – one down staircase procedure: Every incorrect response caused the next pitch difference to be doubled, whereas three successive correct responses led the next pitch difference to be reduced by 50%. The trials stopped after six reversals in the size of the difference. The starting pitch difference was 300 cents (a minor third).

Online Test of Amusia (OTA)

The OTA was constructed to reflect the most critical measures from the full MBEA (Peretz et al., 2003). The OTA serves as an initial screening test for a full evaluation of congenital amusia (Vuvan et al., 2018). The present data therefore cannot be considered as offering a definitive evaluation of this deficit, but may only be said to identify individuals who are potentially amusic. An original version of the OTA was published in 2008 (Peretz et al., 2008) and then later revised for greater validity (Peretz & Vuvan, 2017). We used the revised measure, which comprises three subtests.

The first subtest is a standard/comparison pitch detection task in which participants are presented with thirty trials consisting of two short melodies. Participants are asked to judge if the two melodies are the same or different. On half the trials, the same melody is repeated twice. In the other half, one of the pitches in the second presentation is altered to match a pitch outside the key of the melody, while maintaining the original melodic contour.

The next subtest measures rhythm perception. Participants are presented with twenty-four trials consisting of a single melody and asked to judge whether or not there is an unusual pause in the melody. In half of the trials a 357 millisecond pause delays the timing of one onset, thus disrupting the rhythm. This pause occurs prior to the first downbeat of the third measure.

In the third subtest, participants are presented with twenty-four trials consisting of a single melody and asked to judge if there is an out-of-tune note, which in this subtest is a single pitch that falls outside the tonal context of the melody. Half the trials in this test include a single mistuned note. The critical difference between subtests 1 and 3 is that participants cannot rely on short-term memory for a specific melody to complete trials in subtest 3 but instead must rely on the application of learned tonal schemata.

Procedure

Participants were invited into the lab and were read an informed consent document. Once their consent was obtained, the participant was seated in front of a computer screen. Participants were then outfitted with the headset microphone and headphones for recording and stimulus presentation, respectively. The experimenter controlled both measures at a separate computer monitor and read the on-screen instructions aloud. The tests were counterbalanced with half the participants starting with the SSAP and the other half starting with the OTA. The score output for both measures was hidden from the participants during the experiment in an effort to reduce possible bias on the subsequent test. Both tests took, on average, 45 minutes to complete, about 30 minutes for the OTA and about 15 minutes for the SSAP.

Data Analysis

Accuracy of vocal pitch imitation (SSAP) was based on an automated procedure performed in Matlab (Mathworks, Natick, MA). The fundamental frequency (F0) of produced vocal pitch was extracted using the YIN algorithm (de Cheveigné & Kawahara, 2002). Individual notes for melody imitation trials were then segmented based on amplitude fluctuations associated with the onset of the syllable "doo". The median for the middle 50% of samples associated with an individual note was then used as an estimate of sung pitch. When this estimate fell within ± 50 cents of the target pitch, the sung pitch was coded as accurate, otherwise it was coded as an error. The proportion of accurately produced pitches across all sung pitches was used as a measure of *singing accuracy* for a participant (for further details, see Pfordresher & Demorest, in press).

Pitch discrimination ability (SSAP) was expressed in cents (100 cents = 1 semitone) and was based on the last pitch difference the participant discriminated. This constituted our measure for the perception of *psychophysical pitch discrimination*. Problematic discrimination thresholds were defined as those equal to or greater than 300 cents. Because the starting pitch difference was 300 cents, obtained thresholds greater than this value suggest that participants may have misunderstood the judgment task, mislabeling "higher" versus "lower" pitch changes, and warranted discarding of that participant's data.

Accuracy on the OTA was based on the proportion of correct responses in each subtask. Because the main focus of this paper is on concordance between the OTA and accuracy of sung pitch, we aggregated across the first and third subtests, but excluding the second (rhythmic) subtest of the OTA for a global measure of accuracy in *musical pitch perception*.[1]

Results

During initial screening of the data, we removed from consideration any participant with a pitch discrimination threshold at or greater than the initial pitch difference of 300 cents. Fourteen participants of the initial sample (N = 100) exhibited this behavior and were removed from the sample used for all analyses reported here (N = 86).

We first report bivariate correlations based on global measures of singing accuracy (SSAP), musical pitch perception (OTA), and psychophysical pitch discrimination (SSAP). These are shown in the top row of Figure 2. All correlations were statistically significant, reflecting a greater degree of concordance between singing accuracy and pitch perception than is implied in much of the previous literature.[2] We also examined relationships between each of these measures and years of reported musical training, shown in the bottom row of Figure 2. These correlations were also significant, suggesting that instrumental training is associated with both musical pitch perception, as would be expected, and (less obviously) singing accuracy.

Multiple regression analyses were then used to determine the relative contributions of each performance measure, along with musical training. Three regression analyses were conducted with each primary measure (singing accuracy, musical pitch perception, psychophysical pitch discrimination) functioning as a Y variable. All remaining variables were entered as predictors simultaneously (e.g., regressing singing accuracy simultaneously on musical pitch perception, psychophysical pitch perception, and years of musical training). Standardized regression weights and their statistical significance are shown in Table 1. Both musical pitch perception (global OTA) and psychophysical pitch discrimination thresholds predicted a significant proportion of variance in singing accuracy (SSAP) independently, but musical training did not. Thus, perception and production of pitch may rely on resources that exist independent of formal instruction.

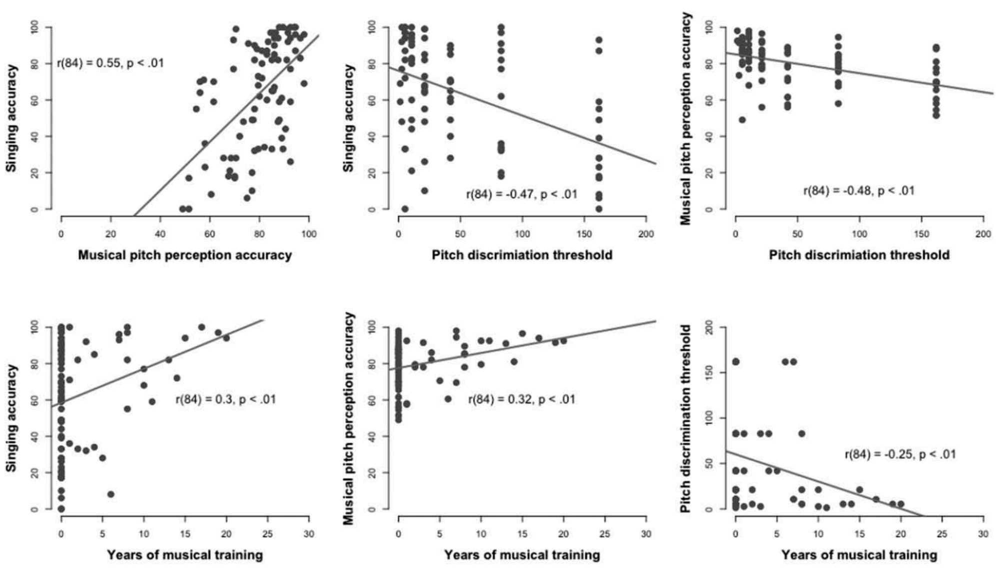

Figure 2. Bivariate relationships for measures of perception, production, and musical background.

Table 1. Standardized regression coefficients and statistical significance.

	Y Variable		
Predictor	Singing accuracy	Musical pitch	Pitch discrimination
Singing accuracy		0.39***	−0.30*
Musical pitch perception	0.40***		−0.28*
Pitch discrimination	−0.25*	−0.26*	
Years of musical training	0.11	0.14	−0.07
Multiple *R*^2	0.38***	0.29***	0.30***

**p*< .05, ****p* < .001

Other regressions yielded similar results, with musical training failing to predict a significant independent portion of the variance in each Y variable.[3]

Further exploration of the proportion of variance accounted for by each predictor (excluding musical training) are shown in Figure 3. These figures were derived by partitioning variance into unique and non-unique portions, treating each measure as the Y-variable (following Cohen & Cohen, 1983, pp. 88–89). For instance, one can start by regressing singing accuracy on musical pitch perception and psychophysical pitch discrimination to yield a total R^2. Semi-partial correlations of each predictor with singing accuracy estimate the variance in singing accuracy that is uniquely associated with that predictor. The difference between the sum of these semi-partial coefficients and the total R^2 in turn can estimate the non-unique (overlapping) variance in singing accuracy accounted for by both predictors (musical pitch perception and psychophysical pitch discrimination). Estimates were highly stable regardless of which variable

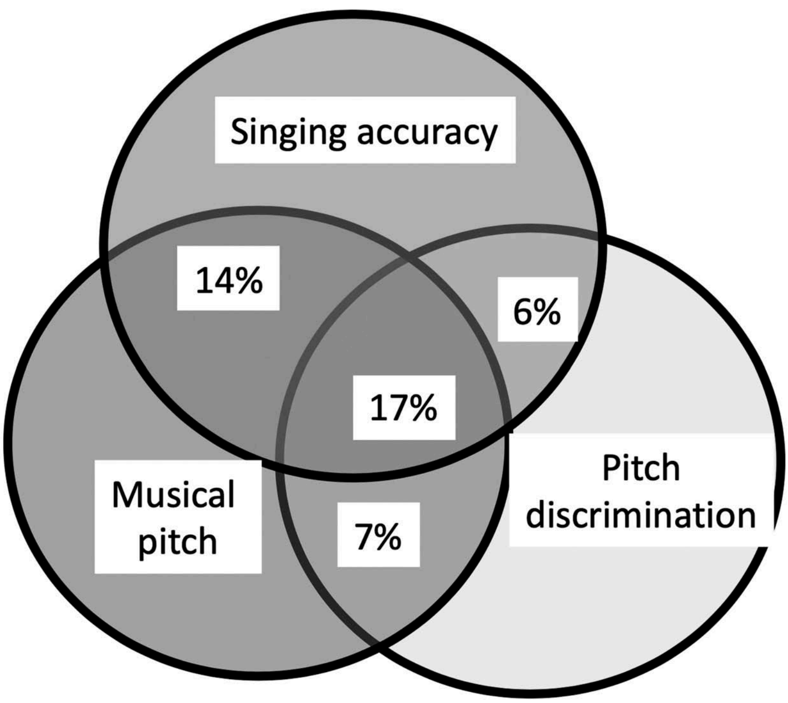

Figure 3. Venn diagram illustrated shared variances across measures of production and perception.

functioned as the outcome (Y). Based on the estimates shown in Figure 3 (derived from on means across 3 regressions), psychophysical pitch discrimination accounts for a much lower portion of unique variance in singing accuracy and musical pitch perception.

Next, we performed a categorical analyses to estimate conditional probabilities pertaining to Congenital Amusia and VPID. Each participant was classified as potentially amusic based on guidelines from Vuvan et al. (2018), in which a score lower than 70% correct on each pitch-based subtest indicates potential amusia (a full diagnosis of which requires validation in subsequent tests which were not used here). Seven participants (about 8% of the sample) were classified as potentially amusic based on this standard (See Table 2). The cutoff for VPID was based on a sample of 632 participants from the SSAP database whose data had been verified for validity (e.g., no recording errors). Participants with accuracy scores more than 2 standard deviations below the mean of this sample (28% correct cutoff) were classified as VPID. Thirteen participants (about 15% of the sample) were classified as VPID based on this criterion. Although this classification strategy differs from strategies used in the past, based on a-priori thresholds (e.g., Berkowska & Dalla Bella, 2009; Pfordresher & Brown, 2007), there is as yet no firmly established practice for differentiating accurate versus poor-pitch singers (Pfordresher & Larrouy-Maestri, 2015). Moreover, we considered the use of statistically-based cutoffs for each test to reflect the most balanced approach to comparing potential deficits given that the practice advocated by Vuvan et al. (2018) were derived from previous studies that used cutoffs based on deviations from the mean (e.g., Henry & McAuley, 2013; Peretz et al., 2003; Peretz & Vuvan, 2017; Pfeifer & Hamann, 2015).

We adopted a conditional probability analysis to determine whether classification with respect to one deficit (VPID or amusic), predicts classification with respect to the other. This analysis is similar to correlation in that it addresses degree of association, but based on categorization rather than scoring individuals on a continuum. Table 2 displays cross-tabulations based on classification of each participant as potentially amusic and/or VPID.

We first consider the conditional probability that an individual is classified as VPID given an existing classification of amusic, *p* (VPID | amusic). According to classic probability theory this can be computed using the following formula:

$$p(\text{VPID} \mid \text{amusic}) \ = p(\text{amusic} \,\&\, \text{VPID}) \ / p(\text{amusic})$$

According to Table 2, the probability of classification as both amusic and VPID, *p* (amusic & VPID) is approximately 5/86 = .06, and the probability of classification as amusic regardless of VPID classification, *p*(amusic), is approximately 7/86 = .08. As

Table 2. Cross-tabulation contingency table representing classification as potentially amusic and VPID.

	Potentially amusic?		
	No	*Yes*	*SUM*
VPID?			
No	71	2	73
Yes	8	5	13
SUM	79	7	

a result, the probability of VPID classification for those already classified as potentially amusic is approximately .06/.08 = .75.

The inverse classification, probability of amusia given classification as VPID, is computed similarly:

$$p(\text{amusic} \mid \text{VPID}) \ = p(\text{amusic} \,\&\, \text{VPID}) \ / p(\text{VPID})$$

Based on the data in Table 2, this probability is approximately .06/.15 = .4; considerably lower than the probability of VPID given amusia. Therefore, in line with the predictions of the VSL model, the presence of amusia is a better predictor of VPID than the reverse. In other words, amusia may be likely to lead to downstream deficits in vocal pitch imitation.

In addition to these observations, the classification results add additional support for strong concordance between these deficits. The probability of classification as both VPID and amusic (.06) was far greater than the rate that would be predicted by independence, p(VPID) * p(amusic) = .01.

Finally, we analyzed associations between all individual subtests within the SSAP and OTA. These subtests are designed to identify different subtypes within those who exhibit a deficiency of perception or production (Berkowska & Dalla Bella, 2013; Pfeifer & Hamann, 2015). Each individual vocal imitation measure in the SSAP was regressed on all three OTA subtests and psychophysical pitch discrimination. Likewise, each OTA subtest was regressed on each SSAP subtest (imitation and pitch discrimination). For these analyses, we included the rhythm subtest of the OTA, which was excluded from earlier analyses on global scores. For purposes of statistical validity, we addressed the fact that measures of pitch discrimination took part in 6 different regressions, whereas other subtests took part in 3 regressions each. Erring on the side of conservatism, a Bonferroni correction was applied leading to $\alpha = .008$ per comparison, based on 6 familywise tests while assessing significance of each predictor. Figure 4 displays significant partial correlation values across all regressions, with arrows pointing from significant predictors to Y-variables, with details in (see Table A1).

This analysis suggests that the broad convergence seen in global SSAP and OTA scores was driven by a smaller number of sub-processes. Most critically, although singing accuracy exhibited broad convergence with pitch measures in general, the only SSAP subtest that correlated with other subtests was the accuracy of imitation of single vocal pitches. Furthermore, the only OTA subtest that predicted single vocal-pitch matching was the third subtest, in which participants detected mistuned notes within a melody. No SSAP imitation subtests regressed significantly onto OTA subtests, consistent with the implications of the classification results. Finally, pitch discrimination measures correlated with only one subtest from the OTA and SSAP.

Discussion

In this study, participants completed two standardized measures. One measure was used for the purpose of testing singing accuracy and the assessment of VPID (SSAP), and the other one was used for the purpose of testing musical perception ability and the assessment of amusia (OTA). Furthermore, psychophysical pitch discrimination thresholds were assessed using a subtest of the SSAP.

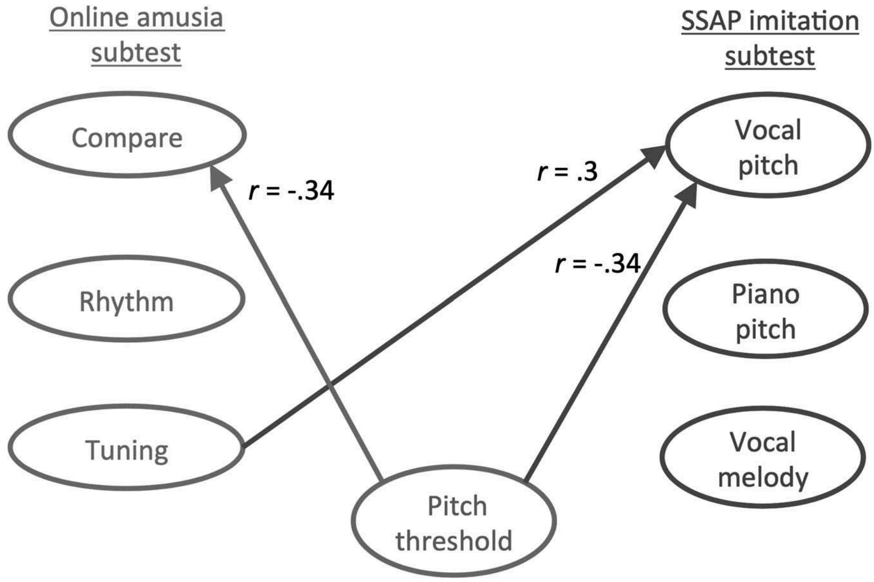

Figure 4. Associations among SSAP and OTA subtests, along with psychophysical pitch discrimination. Numbers show partial regression coefficients for significant predictors.

The results showed a strong positive correlation between the imitation tasks of the SSAP and the pitch trials of the OTA. Participant categorization analysis likewise revealed higher concordance of individuals classified as potentially amusic and VPID than would be predicted by independence. These results contradict several past studies that have focused on dissociations, discussed in the introduction. Recall, however, that these studies did not use the two standardized measures applied here, and the one (unpublished) study that most closely approximates the present research did find a significant association (Wise, 2009). Thus, congenital amusia and VPID may be largely shared deficits, as predicted by the VSL. Also in keeping with this model, participants classified as potentially amusic were much more likely to also have classification as VPID than the reverse.

Another result that informs this outcome arose from relatively weak associations between psychophysical pitch discrimination and either singing accuracy or music perception (Figure 3). Whereas the VSL would predict that pitch discrimination takes prominence over other processes, multiple regression analyses on global scores suggest that pitch discrimination accounted for relatively little unique variance in either singing accuracy or musical pitch discrimination. Further analyses suggested that pitch discrimination may only predict select sub-processes, such as the accuracy of singing single pitches produced with a vocal timbre. Within the present correlational designs any conclusions about the source of these results must be tentative. Nevertheless, it appears as though performance on the OTA and the SSAP may draw on a common higher-order representation of musical structure that may not be necessary for simple pitch discrimination. Importantly, previous dissociations between singing and pitch perception may therefore reflect comparisons between singing and simple psychophysical pitch discrimination (Loui et al., 2008; Pfordresher & Brown, 2007).

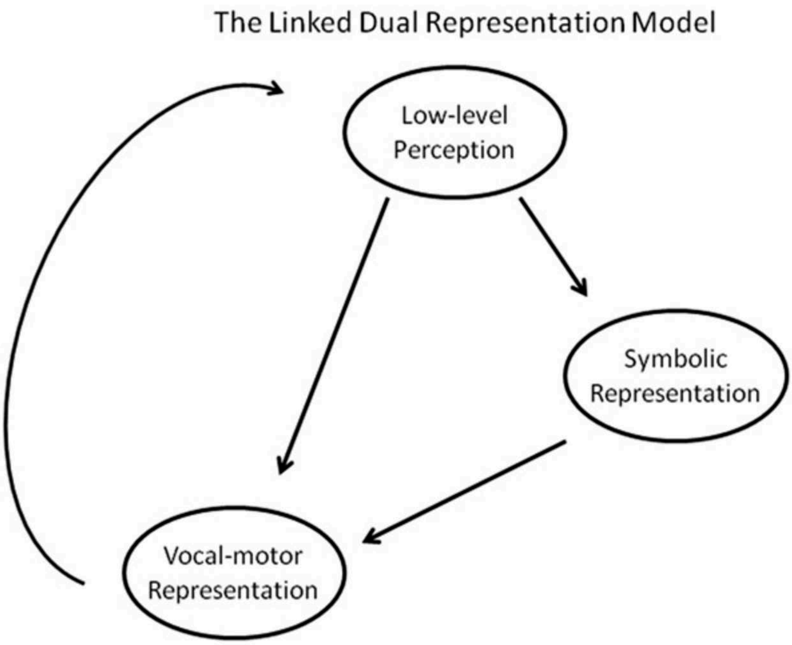

Figure 5. The linked-dual representation model (Hutchins & Moreno, 2013).

Theoretically, these results are consistent with the Linked Dual Representation model, shown in Figure 5 (Hutchins & Moreno, 2013). This model proposes that a mediating symbolic representation may contribute to singing accuracy. Importantly, musical pitch perception (as in the OTA) and singing accuracy (particularly of melodies with tonal relationships) may both rely on this symbolic representation, whereas low-level pitch discrimination does not. This model accounts for the significant proportion of shared variance between singing accuracy and musical pitch perception that is independent of pitch discrimination. The low contribution of pitch perception, however, is still difficult to resolve with this model.

A third goal of this study was to explore whether classification as potentially amusic predicted classification as VPID, as predicted by both the VSL and the Linked Dual Representation model. Classification results supported this hypothesis. Although the categorization analysis is thought-provoking, these results needed to be treated cautiously at present. As noted earlier, our participant categorization analysis was based on cutoffs that were derived in similar ways across SSAP and OTA measures. We did this in order to account for the fact that these measures, and the underlying behaviors, may scale differently, thus warranting the use of standardized cutoffs based on norming samples. There are deep questions associated with the determination of cutoffs for VPID (Pfordresher & Larrouy-Maestri, 2015) and congenital amusia (Henry & McAuley, 2010) that bear on the present data. For this reason, we consider it important that the classification analysis is complemented by regression analyses that analyze performance on a continuum.

The percent of individuals classified as potentially amusic in our sample, 8%, was higher than reported in many other papers. A recent estimate of congenital amusia in the general population, based on the OTA, reports a prevalence of 1.5% (Peretz & Vuvan, 2017), whereas earlier reports (not using the OTA) estimated the frequency at 4% (Kalmus & Fry, 1980). It is not clear why our frequencies were so much higher, but one possibility has to do with differences in recruitment. Whereas these previous studies of amusia relied on participants who sought out the opportunity to participate via advertisements or word of mouth, the present participants participated to fulfill a required portion of a college class and thus may have been less motivated to succeed at the task. Consistent with this interpretation, Pfeifer and Hamann (2015) also recruited individuals who participated as part of a course requirement and found that a higher rate of potential amusia (6% or 9% depending on the measure used) than in many other studies. Obviously more research is required to form a definitive link between the motivation for participating and performance on tests of congenital amusia.

In conclusion, these results have important implications for future research, as well as how to interpret past research. First, it is possible that evidence for dissociations between perception and action systems may have been overstated in the past. Although it is true that inaccurate singing can exist with intact basic pitch perception, higher-level pitch processing involved in music perception may involve processes that overlap substantially with singing accuracy. Moreover, small sample sizes may have led to apparent dissociations that do not hold up in the larger present sample. Second, singing and music perception may rely on common abstract representations of musical structure that do not play a critical role in more basic perceptual and motor tasks. Third, deficits associated with music perception may predict concurrent deficits in singing accuracy, as predicted by leading models.

Notes

1. Based on concerns expressed in the literature about the role of response bias in diagnosis of amusia (Henry & McAuley, 2013; Pfeifer & Hamann, 2015), we also analyzed the data using d-prime. This measure correlated highly with percent correct across participants $r = .98$.
2. Correlations were significant and of similar magnitude when the sample was restricted to participants reporting no formal musical training (N = 59): r(singing accuracy, musical pitch) = .48, r(singing accuracy, pitch discrimination) = .43, r(musical pitch, pitch discrimination) = −.39.
3. When the 14 participants with questionable pitch discrimination thresholds were included, no associations with pitch discrimination were significant.

Disclosure statement

No potential conflict of interest was reported by the authors.

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.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69–83.

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The sung performance battery (SPB). *Frontiers in Psychology*, *4*, 714.

Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2010). Choral singing and psychological wellbeing: Quantitative and qualitative findings from English choirs in a cross-national survey. *Journal of Applied Arts and Health*, *1*, 19–34.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed. ed.). Hillsdale, NJ: Lawrence Earlbaum.

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

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America*, *121*, 1182–1189.

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America*, *126*, 414–424.

de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, *111*, 1917–1931.

Demorest, S. M. (2001). Pitch-matching performance of junior high boys: A comparison of perception and production. *Bulletin of the Council for Research in Music Education*, *151*, 63–70.

Demorest, S. M., & Pfordresher, P. Q. (2015). Seattle singing accuracy protocol - SSAP [measurement instrument]. Retrieved from https://ssap.music.northwestern.edu/.

Demorest, S. M., Pfordresher, P. Q., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: An introduction to the special issue on singing accuracy (Part 2). *Music Perception*, *32*, 266–271.

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception*, *34*, 585–604.

Henry, M. J., & McAuley, J. D. (2010). On the prevalanece of congenital amusia. *Music Perception*, *27*, 413–424.

Henry, M. J., & McAuley, J. D. (2013). Failure to apply signal detection theory to the Montreal battery of evaluation of Amusia may misdiagnose amusia. *Music Perception: An Interdisciplinary Journal*, *30*, 480–496.

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception & Psychophysics*, *76*, 2522–2530.

Hutchins, S., & Moreno, S. (2013). The linked dual representation model of vocal perception and production. *Frontiers in Psychology*, *4*, 825.

Hutchins, S., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, *141*, 76–97.

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

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics*, *43*, 369–382.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*, R331–332.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of speech and song. *Cognition*, *127*, 177–202.

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences*, *20*, 857–867.

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*, 185.

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*, 58–75.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Babnon, B., Trimmer, C. G., … Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception*, *25*, 331–343.

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

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.
Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of “tone deafness”. *Music Perception*, *25*, 95–115.
Pfordresher, P. Q., & Demorest, S. M. (in press). Construction and validation of the seattle singing accuracy protocol (SSAP): An automated online measure of singing accuracy. In F. Russo, B. Ilari, & A. Cohen (Eds.), *Routledge companion to interdisciplinary studies in singing: Vol. 1 development*. London: Routledge.
Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we understand deficits of vocal pitch imitation? *Frontiers in Human Neuroscience*, *9*, 271.
Stewart, N. A. J., & Lonsdale, A. J. (2016). It’s better together: The psychological benefits of singing in a choir. *Psychology of Music*, *44*, 1240–1254.
Vuvan, D. T., Paquette, S., Mignault Goulet, G., Royal, I., Felezeu, M., & Peretz, I. (2018). The Montreal protocol for identification of Amusia. *Behavior Research Methods*, *50*, 662–672.
Welch, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music*, *7*, 50–58.
Welch, G. F. (2005). Singing as communication. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical communication* (pp. 239–259). New York: Oxford University Press.
Welch, G. F., Himonides, E., Saunders, J., Papageorgi, I., & Sarazin, M. (2014). Singing and social inclusion. *Frontiers in Psychology*, *5*.
Williamson, V. J., Liu, F., Peryer, G., Grierson, M., & Stewart, L. (2012). Perception and action de-coupling in congenital amusia: Sensitivity to task demands. *Neuropsychologia*, *50*, 172–180.
Wise, K. (2009). *Understanding “tone deafness”: A multi-componential analysis of perception, cognition, singing and self-perceptions in adults reporting musical difficulties*. (Ph.D.), Keele University, Keele.
Wisniewski, M. G., Mantell, J. T., & Pfordresher, P. Q. (2013). Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind and Brain*, *23*, 82–99.

Appendix A

Table A1. Partial correlations and statistical significance for regressions based on subtests of the SSAP and OTA.

	Y variable					
	Seattle Singing Accuracy Protocol			*Online Test of Amusia*		
Predictor	Vocal pitch	Piano pitch	Vocal melody	Comparison	Rhythm	Tuning
SSAP: Vocal pitch				0.01	−0.06	0.09
SSAP: Piano pitch				0.09	0.14	0.16
SSAP: Vocal melody				0.17	−0.04	0.04
SSAP: Discrimination	−0.33*	−0.24	−0.16	−0.34*	−0.16	−0.18
OTA: Comparison	0.15	0.16	0.24			
OTA: Rhythm	−0.13	−0.06	−0.14			
OTA: Turning	0.30*	0.29	0.21			
Multiple *R*^2	0.36*	0.32*	0.25*	0.32*	0.07	0.27*

*p < .008 (Bonferroni correction)