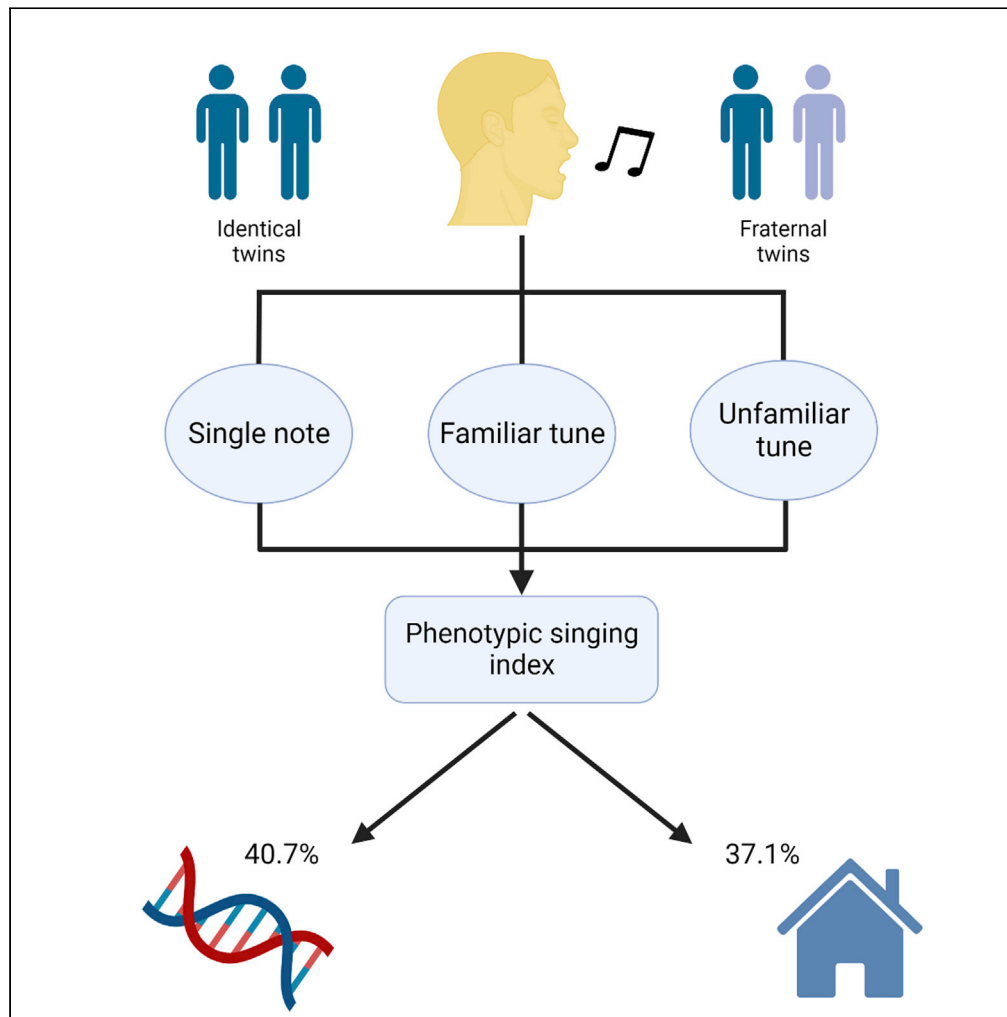


Article

Genetic factors and shared environment contribute equally to objective singing ability



Daniel Yeom, Yi Ting Tan, Nick Haslam, ..., Gary E. McPherson, Isabelle Peretz, Sarah J. Wilson

yeom.d@unimelb.edu.au

Highlights

We measured singing ability in a large sample of Australian twins

Singing ability is moderately heritable

Shared environmental factors are equally important

Of these, early but not current musical environments shaped singing ability

Yeom et al., iScience 25, 104360
June 17, 2022 © 2022 The Authors.
<https://doi.org/10.1016/j.isci.2022.104360>

Article

Genetic factors and shared environment contribute equally to objective singing ability

Daniel Yeom,^{1,8,*} Yi Ting Tan,^{1,2} Nick Haslam,¹ Miriam A. Mosing,^{1,3,4} Valerie M.Z. Yap,¹ Trisnasari Fraser,^{1,2} Michael S. Hildebrand,^{5,6} Sam F. Berkovic,⁵ Gary E. McPherson,² Isabelle Peretz,^{1,7} and Sarah J. Wilson^{1,5}

SUMMARY

Singing ability is a complex human skill influenced by genetic and environmental factors, the relative contributions of which remain unknown. Currently, genetically informative studies using objective measures of singing ability across a range of tasks are limited. We administered a validated online singing tool to measure performance across three everyday singing tasks in Australian twins (n = 1189) to explore the relative genetic and environmental influences on singing ability. We derived a reproducible phenotypic index for singing ability across five performance measures of pitch and interval accuracy. Using this index we found moderate heritability of singing ability ($h^2 = 40.7\%$) with a striking, similar contribution from shared environmental factors ($c^2 = 37.1\%$). Childhood singing in the family home and being surrounded by music early in life both significantly predicted the phenotypic index. Taken together, these findings show that singing ability is equally influenced by genetic and shared environmental factors.

INTRODUCTION

Singing is central to being human, maintaining our emotional wellbeing and sociocultural identity. There is a large amount of variability in musical talent and proficiency, ranging from people with congenital amusia to prodigies. For decades, the model of 10,000 h of deliberate practice has been the accepted explanation for the acquisition and maintenance of music expertise (Ericsson et al., 1993). This model rejects the role of “innate” predispositions on expert development (Ericsson et al., 1993). However, recent work indicates that deliberate practice likely accounts for only 30% of the variance in expert musicianship, implicating other factors (Hambrick et al., 2018; Macnamara et al., 2014). Research has now turned to exploring the behavioral and molecular genetic bases of music ability to assess the unique and interactive effects of genes and environment.

Singing ability is a complex behavioral trait that requires robust, objective phenotyping to establish its genetic basis (Gingras et al., 2015). Singing provides an ideal phenotype for exploring the genetic basis of musical talent as it exists in all cultures, has universal features (Jacoby et al., 2019; Mehr et al., 2019), and emerges early and spontaneously during human development (Welch, 2015). Phenotyping singing ability, however, is not a trivial issue. As singing is a complex behavioral skill involving the coordination of perceptual and productive processes, it is likely influenced by many genetic variants of small effect. Arguably, these could differentially contribute to pitch accuracy, rhythm, timbre, vocal production, lyrical expression, and harmonization with others. The challenge then is to identify the core phenotypic components of singing and their genetic variants.

Limited molecular evidence has identified genetic variants related to music aptitude in humans that are also linked to song learning in songbirds, suggesting shared evolutionary origins for human singing (Järvelä, 2018; Kanduri et al., 2015). The occurrence of untrained individuals with high singing pitch accuracy supports an innate predisposition (Larrouy-Maestri et al., 2013; Watts et al., 2003; Wilson et al., 2011). To date, Park et al. (2012) is the only study to explore the genetic basis of singing ability using an objective measure. They used a simple vocal pitch-matching task in an isolated Mongolian sample showing that vocal pitch-matching is moderately heritable ($h^2 = 40\%$) and linked to *UGT8*. However, their isolated homogeneous sample limited the investigation of the role of environmental influences, and their use of a simple

¹Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, VIC 3010, Australia

²Melbourne Conservatorium of Music, University of Melbourne, Southbank, VIC 3006, Australia

³Department of Neuroscience, Karolinska Institutet, 17177 Stockholm, Sweden

⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 17177 Stockholm, Sweden

⁵Department of Medicine, Epilepsy Research Centre, University of Melbourne, Austin Health, Heidelberg, VIC 3084, Australia

⁶Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, VIC 3052, Australia

⁷International Laboratory for Brain, Music and Sound Research and Department of Psychology, University of Montreal, Montreal, QC H2V 2S9, Canada

⁸Lead contact

*Correspondence: yeom.d@unimelb.edu.au
<https://doi.org/10.1016/j.isci.2022.104360>



Table 1. Summary of the three singing tasks in the Melbourne Singing Tool

Task	Task requirements	No. trials	Frequency range of stimuli	Variable extracted
Single Note	Sing back 5 sine tones Pitch classes B, C#, D#, F, G	10	Female: 246.94–397Hz Male: 123.47–196Hz	PD
Familiar Tune	Sing Happy Birthday in different conditions (paced/unpaced, with/without lyrics). The paced condition was used in the current study, sung first with the lyrics and then twice on the syllable “dah” after listening to a drum beat of Happy Birthday at 120bpm ^a	3	Participants choose their preferred starting note	PD ID
Unfamiliar Tune	Sing back 5 different 7-note piano tunes	10	Female: 220–440 Hz Male: 110–220Hz	PD ID

PD, pitch deviation; ID, interval deviation.

^aThe average of trials two and three was used to measure performance as evidence suggests that singing with a syllable produces a more accurate performance than singing with lyrics (Berkowska and Dalla Bella, 2013; Dalla Bella et al., 2007; Tan et al., 2021).

pitch-matching task did not sufficiently capture the range of skills required for everyday singing. Two additional studies (Coon and Carey, 1989; Morley et al., 2012) have also investigated the genetic basis of participating in singing, which may not be related to objective singing ability. There is a critical need to expand on previous work using more robust and comprehensive measures that capture the complexity of singing phenotypes.

Here, we administered a recently validated online singing tool, the Melbourne Singing Tool (Tan et al., 2021), in a large sample of Australian twins ($n = 1189$) to objectively measure singing pitch accuracy. We hypothesized that singing accuracy would show evidence of heritability. As a secondary aim, we explored the relationship between singing pitch accuracy performance and environmental variables such as early and current experiences of singing with family.

RESULTS

Objective performance on singing tasks

The Melbourne Singing Tool included three main singing tasks: vocal pitch-matching (Single Note task), singing “Happy Birthday” (Familiar Tune task), and singing novel melodies (Unfamiliar Tune task; see Table 1 and STAR Methods). These tasks were designed to capture “everyday” singing behaviors that the general population typically engages in irrespective of professional singing training. They were also chosen for comparability with tools used in previous studies (Berkowska and Dalla Bella, 2013; Pfordresher and Demorest, 2020).

The tasks yielded five measures of singing pitch accuracy for each participant – three for mean absolute pitch deviation (PD) and two for mean interval deviation (ID) measured in cents (see STAR Methods). Pitch deviation reflects the average difference between the frequency of a sung note and a target note for each task. Interval deviation reflects the average difference between a sung interval (the space between two notes in a tune) and a target interval for the tune tasks, thereby capturing singing accuracy over time. For each of the raw measures, lower deviation values reflect greater accuracy to the target pitch/interval and therefore higher performance.

Performance on all five measures was heavily skewed toward lower deviation values. Median performance on each measure fell below 100 cents (Figure 1A), a deviation of less than one semitone, and a common cut-off score for accurate singing (Dalla Bella and Berkowska, 2009; Dalla Bella et al., 2007). For the tune tasks, the medians for the PD and ID measures were similar for each task. The lowest median PD performance was observed for the Single Note task likely reflecting its simplicity, whereas the Familiar Tune task showed the least variance likely owing to its high familiarity. In general, however, performance across the tasks was similar (Table 2).

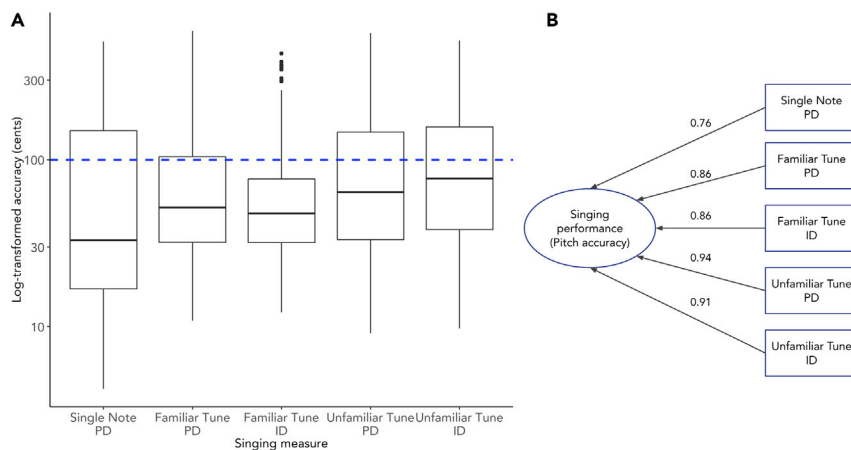


Figure 1. The five singing measures used to construct the singing ability phenotypic index (n = 1189)

(A) Boxplots showing accuracy in cents on the five singing measures, presented on a log-transformed scale. Lower scores represent greater accuracy. The blue dotted line marks a deviation of 100 cents (one semitone).

(B) Loadings of the five singing measures on the single singing performance factor. PD = pitch deviation; ID = interval deviation.

A singing ability phenotypic index

To explore possible latent factors that capture singing ability across tasks we conducted an exploratory factor analysis with principal axis factoring on the five measures. First, we applied a natural log transformation to each measure to minimize the skew from its distribution. For the ease of interpretation of results, we then reverse-coded the log-transformed values so that higher values reflect better singing performance on each measure. The five measures showed high internal reliability ($\alpha = 0.92$), with factor analysis indicating a single factor provided the best solution. All five measures loaded strongly on this factor (Figure 1B) explaining 75.3% of the variance in singing performance. Factor scores for this latent variable were calculated using Thurstone's regression method and are henceforth referred to as the singing factor score. Of note, the factor score correlated highly with the average of the five log-transformed measures ($r = 0.986$, 95% confidence interval (CI): [0.984, 0.987]). The high internal reliability and strong loadings of this factor suggest that it is a robust phenotypic index, capturing the complexity of everyday singing ability.

Heritability of singing ability

To examine whether individual differences in singing ability are influenced by genetic factors, we first calculated twin correlations on the singing factor score, split by zygosity and adjusted for age and sex during modeling (Table 3). This showed similarly high twin correlations for male and female monozygotic (MZ) twins compared to consistently lower twin correlations for dizygotic (DZ) twins, with assumption tests showing no significant sex differences ($p > 0.01$; Table 4). Modeling the combined data we found MZ twins had a higher twin correlation ($r_{MZ} = 0.78$, 95% CI: [0.73, 0.81]) than DZ twins ($r_{DZ} = 0.58$, 95% CI: [0.46, 0.68]) with nonoverlapping CIs (Figure 2).

The twin correlation for MZ twins was less than double that of DZ twins, suggesting that shared familial factors (c^2) rather than dominant genetic factors play a role. We, therefore, used structural equation modeling to examine the influence of shared environmental effects, fitting a univariate ACE model over an ADE model (Grasby et al., 2017). Supporting our main hypothesis, singing ability showed a moderate level of heritability ($h^2 = 40.7\%$, 95% CI: [22.5%, 63.4%]). Strikingly, we found a similar level of influence of shared environmental factors ($c^2 = 37.1\%$, 95% CI: [14.7%, 54.7%]). In contrast, unshared environmental factors accounted for only 22.2% (95% CI: [18.8%, 26.3%]) of the phenotypic variance in singing ability. To assess whether these estimates were significant, we fitted a series of sub-models without the A, C, or E path (Table 5). All sub-models led to a significantly worse fit compared to the full ACE model ($p < 0.01$), and so all paths were retained.

The importance of early shared musical environments

In view of the strong evidence for the role of shared environmental factors in singing ability, we regressed the singing factor score against three self-report measures of childhood and current shared musical

Table 2. Descriptive statistics for the five singing measures in cents in the twin sample (n = 1189)

Singing measure	M	SD	Participants scoring <100 cents	Median
Single Note PD	96.99	118.47	69.0%	32.92
Familiar Tune PD	81.35	77.42	73.4%	51.83
Familiar Tune ID	61.91	46.80	86.0%	47.65
Unfamiliar Tune PD	114.66	119.31	64.1%	63.97
Unfamiliar Tune ID	108.70	90.59	58.5%	77.21

PD, pitch deviation; ID, interval deviation.

experiences from the Melbourne Singing Tool (n = 1174; see [STAR Methods](#)). We chose measures that likely captured early environmental influences that twin pairs would experience together, as well as a related measure of current shared environmental influence. These included early experiences of singing with family and being surrounded by music in childhood, as well as current singing with family. We included age and sex as covariates in our analysis.

Singing with family and being surrounded by music in childhood predicted a higher singing factor score, and therefore more accurate performance ($p < 0.001$). In contrast, current singing with family did not predict singing performance ($p = 0.852$). This model explained 26.5% of the variance, and age was a significant negative covariate ($p < 0.001$; [Table 6](#)).

DISCUSSION

Using a well-validated multi-platform online singing tool, we derived a phenotypic index of singing ability that objectively quantitates pitch accuracy for a range of everyday singing tasks. Importantly, we showed that this phenotypic index is heritable, at a level consistent with previous behavioral studies of >17,000 human traits ([Polderman et al., 2015](#)) including music perception and production ([Drayna et al., 2001](#); [Mosing et al., 2014b](#); [Ullén et al., 2014](#)). Strikingly, we also demonstrated an equal impact of the shared environment on this phenotypic index, which was more than double the average effect reported in previous behavioral genetics studies ([Polderman et al., 2015](#)). Further analyses provided clues for the type of environmental experiences that may shape singing ability, including early singing with family and being surrounded by music in childhood. These findings point to a combined equal effect of genetic and environmental influences in determining an individual's objective singing performance.

We used tasks that the general population routinely engage in, such as singing well-known tunes on special occasions and learning new tunes through imitation. Our sample's performance on each of the five measures derived from the Melbourne Singing Tool is commensurate with the general population as reported in behavioral singing research ([Dalla Bella and Berkowska, 2009](#); [Pfordresher and Brown, 2007](#); [Pfordresher et al., 2010](#); [Pfordresher and Demorest, 2021](#); [Tan et al., 2021](#)). Although there was substantial variation in performance between individuals, the distribution of singing accuracy on each of the measures was skewed toward accurate performance. Most participants demonstrated a moderate level of accuracy, with the majority singing target pitches or pitch intervals, on average within one semitone ([Dalla Bella and Berkowska, 2009](#); [Pfordresher and Brown, 2007](#); [Pfordresher et al., 2010](#); [Pfordresher and Demorest,](#)

Table 3. Twin correlations for the singing factor score by zygosity (n = 517 twin pairs)

Zygosity group	N pairs (% ^a)	Twin correlation [95% CI]
MZ male	98 (26%)	0.75 [0.65, 0.83]
MZ female	279 (74%)	0.78 [0.73, 0.83]
DZ male	30 (23%)	0.50 [0.18, 0.72]
DZ female	99 (77%)	0.60 [0.46, 0.71]
DZ opposite sex	11	0.43 [-0.15, 0.79]

MZ, monozygotic; DZ, dizygotic.

^a% for same sex zygotic twin pairs.

Table 4. Model-fitting statistics for the assumption testing of quantitative and qualitative sex differences

Model	Base model	Comparison model	Estimated parameters	-2LL	AIC	df	Change in -2LL	Change in df	p
1	Saturated model		27	2,466.51	452.51	1007	NA	NA	NA
2	Saturated model	Means equated across twin order	23	2,467.38	445.38	1011	0.864	4	0.930
3	Means equated across twin order	Means equated across zygosity	21	2,468.61	442.61	1013	1.231	2	0.541
4	Means equated across zygosity	Means across same sex	19	2,476.11	446.11	1015	7.501	2	0.024
5	Means across same sex	Means across both sexes	18	2,476.11	444.11	1016	<0.001	1	1
6	Means across both sexes	Variances equated across twin order	14	2,477.62	437.62	1020	1.514	4	0.824
7	Variances equated across twin order	Variances equated across zygosity	12	2,477.64	433.64	1022	0.02	2	0.990
8	Variances equated across zygosity	Variances equated across same sex	10	2,478.30	430.30	1024	0.656	2	0.720
9	Variances equated across same sex	Variances equated across both sexes	9	2,480.41	430.41	1025	2.111	1	0.146
10	Variances equated across both sexes	Covariances equated between MZ male and female	8	2,480.41	428.41	1026	0.004	1	0.951
11	Covariances equated between MZ male and female	Covariances equated between DZ male and female	7	2,480.64	426.64	1027	0.229	1	0.632
12	Covariances equated between DZ male and female	Covariances equated across DZ opposite sex	6	2,480.67	424.67	1028	0.028	1	0.868

Alpha criterion is set at $p < 0.01$. -2LL = minus two log-likelihood, AIC = Akaike Information Criterion, df = degrees of freedom. A lower AIC represents a better model fit. A significant p-value ($p < 0.01$) indicates a significantly worse model fit.

2021). This is broadly consistent with previous studies showing that the general population tend to sing accurately (Dalla Bella and Berkowska, 2009; Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Pfordresher et al., 2010; Pfordresher and Demorest, 2021). Our findings were also consistent with recent findings showing a negative relationship between age and accuracy in singing tasks (Pfordresher and Demorest, 2021). Our study shows that objectively measured singing ability can be captured by a strong single factor, allowing us to construct a phenotypic index of everyday singing.

For the first time, we showed that singing ability is heritable to a degree consistent with music genetics and behavioral genetics research (Butkovic et al., 2015; Hambrick and Tucker-Drob, 2015; Mosing et al., 2014b; Polderman et al., 2015; Ullén et al., 2014). In a recent meta-analysis of behavioral genetics studies spanning 50 years, Polderman et al. (2015) showed that the average heritability of human traits is approximately 49%. For musical traits, perceptual processes such as pitch perception ($h^2 = 38\text{--}51\%$) (Mosing et al., 2015; Seesjärvi et al., 2016; Ullén et al., 2014), and tune recognition ($h^2 = 59\text{--}80\%$) (Drayna et al., 2001; Mosing et al., 2014b; Ullén et al., 2014) are also moderately heritable, along with the propensity to practice ($h^2 = 38\text{--}70\%$) (Butkovic et al., 2015; Hambrick and Tucker-Drob, 2015; Mosing et al., 2014a) and attain music achievements ($h^2 = 55\text{--}57\%$ in males, $9\text{--}13\%$ in females) (Mosing et al., 2015; Wesseldijk et al., 2019). Notably, our heritability estimate is consistent with the sole previous study using vocal pitch-matching that reported 40% heritability (Park et al., 2012). To our knowledge, our study is the first to establish the heritability of singing ability using everyday singing tasks such as singing familiar and unfamiliar tunes, adding to the growing body of evidence suggesting that, like many other human traits, musical traits have a genetic basis (Polderman et al., 2015).

A large, shared environment effect is atypical in behavioral genetics research. Many human traits are only weakly influenced by shared environment, if at all (Polderman et al., 2015), with Polderman et al. (2015) finding an average effect of just 17.4%. Our estimate of 37.1% is more than double the average estimates

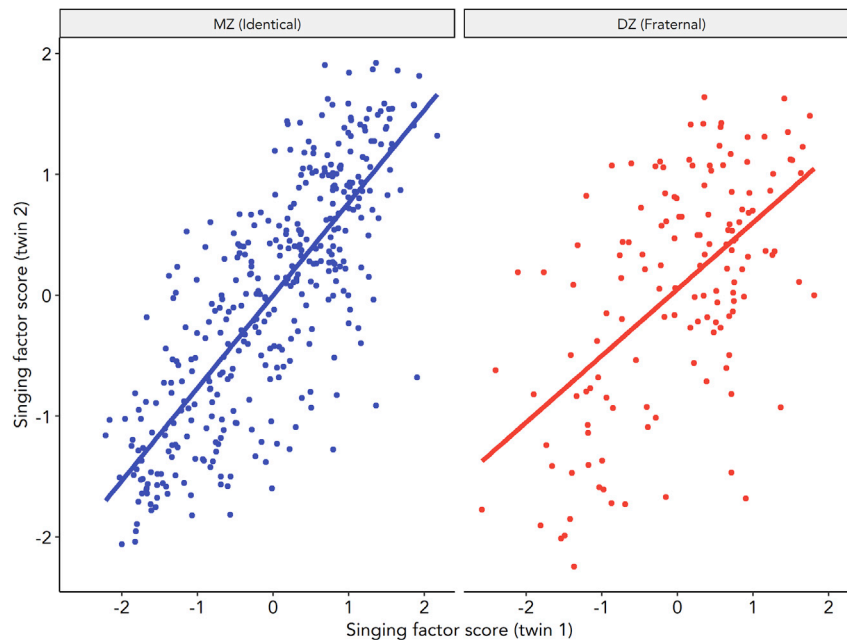


Figure 2. Intra-twin correlations between monozygotic (MZ) and dizygotic (DZ) twins on the singing factor score (n = 1189).

observed in behavioral genetics studies for related phenotypes in cognitive (17.7%) and psychiatric (15.8%) domains, as well as those for social interactions (18.2%) and social values (27.1%) (Polderman et al., 2015). Arguably, this large effect may be partially explained by the importance of early childhood singing and immersion in music for both MZ and DZ twins. It is consistent with evidence for early sensitive periods of music development, where an enriched musical environment is thought to foster neurodevelopmental change supporting music processing (Steele et al., 2013; Vaquero et al., 2016) and the acquisition of musical skills (Hannon and Trehub, 2005; Merrett et al., 2013; Wilson et al., 2012). Evidence also supports a sensitive period for sensorimotor integration, a crucial process that links perception and action for music and singing production (Bailey and Penhune, 2012; Steele et al., 2013) including vocal pitch-matching (Hutchins et al., 2014; Pfordresher and Brown, 2007).

Importantly, Wesseldijk et al. (2019) showed that measures of a musically-enriched childhood correlate with achievements in music and that heritability of musical achievement increases with higher mean levels of musical enrichment. Relatedly, Hambrick and Tucker-Drob (2015) found that the heritability of accomplishment was greater in twins who engaged in regular practice, which in turn has a heritable component (Butkovic et al., 2015; Hambrick and Tucker-Drob, 2015; Mosing et al., 2014a). In the case of pitch-accurate singing, our findings suggest that a genetic predisposition may be more likely expressed when an individual is exposed to an early enriched singing and musical environment, potentially leading to changes in the neural networks underpinning singing (Wilson et al., 2011) that may occur through interactions between genes and early environments. These gene-environment interactions take place when children with genetic predispositions are actively or passively exposed to environments that foster these predispositions. An interaction between genetic and shared familial influences would inflate the estimate of C, and so may explain our high C value (Grasby et al., 2017).

To reliably establish the genetic and neurobiological bases of music-related traits, objective phenotypes need to be established (Gingras et al., 2015). Our findings provide an important first step toward this endeavor for everyday singing ability. Although genes such as *UGT8* (Park et al., 2012) and *SLC6A4* (Morley et al., 2012) have been implicated in previous studies of vocal pitch-matching and choral participation, respectively, to date no studies have explored genetic variants of more complex objective phenotypes of singing ability. Recent findings have shown shared genes between song learning in songbirds and transcriptome regulation in humans after musical performance and listening, providing potential variants of interest underlying these complex phenotypes (Kanduri et al., 2015; Nair et al., 2019, 2020; Oikkonen et al., 2016).

Table 5. Model-fitting statistics for the univariate ACE model, comparing a full ACE model against various constrained submodels

Base model	Comparison model	Estimated parameters	-2LL	AIC	df	Change in -2LL	Change in df	p
ACE		6	2,480.670	424.670	1028			
ACE	AE	5	2,489.763	431.763	1029	9.093	1	0.003
ACE	CE	5	2,505.025	447.025	1029	24.356	1	<0.001
ACE	E	4	2,886.138	826.138	1030	405.469	2	<0.001

A, additive genetic effects; C, shared environmental effects; E, non-shared environmental effects; -2LL, minus two log-likelihood; AIC, Akaike Information Criterion; df, degrees of freedom.

ACE represents the full model, while AE and CE are submodels that drop the C and A paths, respectively.

A lower AIC represents a better model fit. A significant p-value indicates a significantly worse model fit.

In conclusion, we demonstrate an objective phenotypic index of singing ability that captures performance on a range of everyday singing tasks. This phenotype is moderately heritable, suggesting a genetic basis. However, shared environmental factors equally influence this singing ability phenotype. Our findings highlight the need for further investigation of these heritable and shared environmental effects, including the interplay between our genetic predispositions and the environments in which we grow up.

Limitations of the study

There are some limitations to our study. First, the measures of early shared environment that we used in our exploratory analyses were based on self-report. Although they provided an initial indication of shared environmental factors associated with singing ability, they warrant more detailed investigation using prospective longitudinal studies. These measures themselves may also be influenced by genetic factors, and so multivariate modeling may help unpick these relationships. Second, our study did not directly measure the rhythmic accuracy of singing, although our choice of singing tasks captured various rhythms. Our five measures showed similar pitch and interval accuracy scores despite this rhythmic variation and all loaded strongly onto one factor, suggesting that rhythmic ability was not a major factor in our phenotype. Third, self-selection bias may have occurred in our recruitment as people already interested in music and/or singing may have been more likely to participate. Research has clearly described the cultural and social barriers for Australian boys and men to engage in singing activities (Harrison, 2007; Powell, 2014). In addition, Twins Research Australia, who assisted with recruitment (see STAR Methods), has a higher number of females in its database (Murphy et al., 2019). Both these points likely account for our higher female participation rate. Future investigations in samples with lower rates of musical training, and higher proportions of men, would be valuable. However, given that the performance of our sample is commensurate with that reported in the general population (Berkowska and Dalla Bella, 2013; Dalla Bella and Berkowska, 2009; Dalla Bella et al., 2007) it is unlikely this affected our results.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact

Table 6. Multiple regression showing the predictive effect of the shared environmental measures on the singing factor score, adjusted for age and sex (n = 1174)

Variable	B	β	SE	t	P
Singing with family (childhood)	0.269	0.322	0.028	9.790	<0.001
Being surrounded by music (childhood)	0.205	0.228	0.027	7.669	<0.001
Singing with family (current)	0.005	0.006	0.026	0.186	0.852
Age	-0.005	-0.087	0.002	-3.328	<0.001
Sex	-0.067	-0.030	0.057	-1.180	0.238

B, unstandardized beta coefficient; β , standardized beta coefficient; SE, standard error.

- Materials availability
- Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Participants
- **METHOD DETAILS**
 - The Melbourne Singing Tool
 - Singing variables
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104360>.

ACKNOWLEDGMENTS

We thank all the twins for their participation in the study. We thank Dr. Laura Wesseldijk for reviewing the R code and outputs for the univariate genetic analysis, and for her valuable input on the analyses. We also thank Mark Solly and Oscar Correa for their work in developing the Melbourne Singing Tool. This research was supported by an Australian Research Council Discovery Project Grant (DP170102479) and an Australian Government Research Training Program Scholarship. This research was facilitated through access to Twins Research Australia, a national resource supported by a Centre of Research Excellence Grant (ID: 1079102), from the National Health and Medical Research Council. The graphical abstract was created with BioRender (www.biorender.com).

AUTHOR CONTRIBUTIONS

S.J.W., I.P., Y.T.T., G.E.M., and S.F.B. designed research; Y.T.T. and T.F. performed research; D.Y., S.J.W., N.H., V.M.Z.Y., and M.M. analyzed data; and D.Y., S.J.W., N.H., Y.T.T., G.E.M., I.P., M.M., M.S.H., S.F.B., and V.M.Z.Y. wrote the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way.

Received: December 6, 2021

Revised: February 28, 2022

Accepted: May 2, 2022

Published: June 17, 2022

REFERENCES

- Bailey, J., and Penhune, V.B. (2012). A sensitive period for musical training: contributions of age of onset and cognitive abilities. *Ann. N. Y. Acad. Sci.* 1252, 163–170. <https://doi.org/10.1111/j.1749-6632.2011.06434.x>.
- Berkowska, M., and Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: the sung performance battery (SPB). *Front. Psychol.* 4, 714. <https://doi.org/10.3389/fpsyg.2013.00714>.
- Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., et al. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 76, 306–317. <https://doi.org/10.1007/s11336-010-9200-6>.
- Butkovic, A., Ullén, F., and Mosing, M.A. (2015). Personality related traits as predictors of music practice: underlying environmental and genetic influences. *Pers. Individ. Dif.* 74, 133–138. <https://doi.org/10.1016/j.paid.2014.10.006>.
- Coon, H., and Carey, G. (1989). Genetic and environmental determinants of musical ability in twins. *Behav. Genet.* 19, 183–193. <https://doi.org/10.1007/bf01065903>.
- Dalla Bella, S., and Berkowska, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Ann. N. Y. Acad. Sci.* 1169, 99–107. <https://doi.org/10.1111/j.1749-6632.2009.04558.x>.
- Dalla Bella, S., Deutsch, D., Peretz, I., and Deutsch, D. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am.* 121, 1182–1189. <https://doi.org/10.1121/1.2427111>.
- Drayna, D., Manichaikul, A., Lange, M.d., Snieder, H., and Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science* 291, 1969–1972. <https://doi.org/10.1126/science.291.5510.1969>.
- Ericsson, K.A., Krampe, R.T., and Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. <https://doi.org/10.1037/0033-295x.100.3.363>.
- Fox, J., and Dusa, A. (2019). Polycor: Polychoric and Polyserial Correlations (McMaster University).

<https://cran.r-project.org/web/packages/polycor/index.html>.

- Gingras, B., Honing, H., Peretz, I., Trainor, L.J., and Fisher, S.E. (2015). Defining the biological bases of individual differences in musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140092. <https://doi.org/10.1098/rstb.2014.0092>.
- Grasby, K.L., Verweij, K.J.H., Mosing, M.A., Zietsch, B.P., and Medland, S.E. (2017). Estimating heritability from twin studies. In *Statistical Human Genetics*, R. Elston, ed., pp. 171–194. https://doi.org/10.1007/978-1-4939-7274-6_9.
- Hambrick, D.Z., and Tucker-Drob, E.M. (2015). The genetics of music accomplishment: evidence for gene-environment correlation and interaction. *Psychon. Bull. Rev.* 22, 112–120. <https://doi.org/10.3758/s13423-014-0671-9>.
- Hambrick, D.Z., Burgoyne, A.P., Macnamara, B.N., and Ullen, F. (2018). Toward a multifactorial model of expertise: beyond born versus made. *Ann. N. Y. Acad. Sci.* 1423, 284–295. <https://doi.org/10.1111/nyas.13586>.
- Hannon, E.E., and Trehub, S.E. (2005). Tuning in to musical rhythms: infants learn more readily than adults. *Proc. Natl. Acad. Sci. U S A* 102, 12639–12643. <https://doi.org/10.1073/pnas.0504254102>.
- Harrison, S.D. (2007). A perennial problem in gendered participation in music: what's happening to the boys? *Br. J. Music Educ.* 24, 267–280. <https://doi.org/10.1017/s0265051707007577>.
- Hutchins, S., Larrouy-Maestri, P., and Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Atten. Percept. Psychophys.* 76, 2522–2530. <https://doi.org/10.3758/s13414-014-0732-1>.
- Jacoby, N., Undurraga, E.A., McPherson, M.J., Valdés, J., Ossandón, T., and McDermott, J.H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Curr. Biol.* 29, 3229–3243.e12. <https://doi.org/10.1016/j.cub.2019.08.020>.
- Järvelä, I. (2018). Genomics studies on musical aptitude, music perception, and practice. *Ann. N. Y. Acad. Sci.* 1423, 82–91. <https://doi.org/10.1111/nyas.13620>.
- Kanduri, C., Kuusi, T., Ahvenainen, M., Philips, A.K., Lahdesmaki, H., and Järvelä, I. (2015). The effect of music performance on the transcriptome of professional musicians. *Sci. Rep.* 5, 9506. <https://doi.org/10.1038/srep09506>.
- Larrouy-Maestri, P., Leveque, Y., Schon, D., Giovanni, A., and Morsomme, D. (2013). The evaluation of singing voice accuracy: a comparison between subjective and objective methods. *J. Voice* 27, 259.e1–259.e5. <https://doi.org/10.1016/j.jvoice.2012.11.003>.
- Macnamara, B.N., Hambrick, D.Z., and Oswald, F.L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol. Sci.* 25, 1608–1618. <https://doi.org/10.1177/0956797614535810>.
- Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., Bello, J., and Dixon, S. (2015). Computer-aided melody note transcription using the TONY software: accuracy and efficiency. In *Proceedings of the First International Conference on Technologies for Music Notation and Representation*, M. Battier, J. Bresson, and Couprie, C. Davy-Rigaux, D. Fober, Y. Geslin, H. Genevois, F. Picard, and A. Tacaille, eds. (Université Paris-Sorbonne), pp. 23–31.
- Mehr, S.A., Singh, M., Knox, D., Ketter, D.M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A.A., Hopkins, E.J., et al. (2019). Universality and diversity in human song. *Science* 366, eaax0868. <https://doi.org/10.1126/science.aax0868>.
- Merrett, D.L., Peretz, I., and Wilson, S.J. (2013). Moderating variables of music training-induced neuroplasticity: a review and discussion. *Front. Psychol.* 4, 606. <https://doi.org/10.3389/fpsyg.2013.00606>.
- Morley, A.P., Narayanan, M., Mines, R., Molokhia, A., Baxter, S., Craig, G., Lewis, C.M., and Craig, I. (2012). AVPR1A and SLC6A4 polymorphisms in choral singers and non-musicians: a gene association study. *PLoS One* 7, e31763. <https://doi.org/10.1371/journal.pone.0031763>.
- Mosing, M.A., Madison, G., Pedersen, N.L., Kujala-Halkola, R., and Ullen, F. (2014a). Practice does not make perfect: no causal effect of music practice on music ability. *Psychol. Sci.* 25, 1795–1803. <https://doi.org/10.1177/0956797614541990>.
- Mosing, M.A., Pedersen, N.L., Madison, G., and Ullen, F. (2014b). Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. *PLoS One* 9, e113874. <https://doi.org/10.1371/journal.pone.0113874>.
- Mosing, M.A., Verweij, K.J.H., Madison, G., Pedersen, N.L., Zietsch, B.P., and Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evol. Hum. Behav.* 36, 359–366. <https://doi.org/10.1016/j.evolhumbehav.2015.02.004>.
- Murphy, K., Lam, J., Cutler, T., Tyler, J., Calais-Ferreira, L., Li, S., Little, C., Ferreira, P., Craig, J.M., Scurrah, K.J., and Hopper, J.L. (2019). Twins Research Australia: a new paradigm for driving twin research. *Twin Res. Hum. Genet.* 22, 438–445. <https://doi.org/10.1017/thg.2019.101>.
- Nair, P.S., Kuusi, T., Ahvenainen, M., Philips, A.K., and Järvelä, I. (2019). Music-performance regulates microRNAs in professional musicians. *PeerJ* 7, e6660. <https://doi.org/10.7717/peerj.6660>.
- Nair, P.S., Raijas, P., Ahvenainen, M., Philips, A.K., Ukkola-Vuoti, L., and Järvelä, I. (2020). Music-listening regulates human microRNA expression. *Epigenetics* 16, 554–566. <https://doi.org/10.1080/15592294.2020.1809853>.
- Neale, M.C., Hunter, M.D., Pritikin, J.N., Zahery, M., Brick, T.R., Kirkpatrick, R.M., Estabrook, R., Bates, T.C., Maes, H.H., and Boker, S.M. (2016). OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika* 81, 535–549. <https://doi.org/10.1007/s11336-014-9435-8>.
- Oikkonen, J., Kuusi, T., Peltonen, P., Raijas, P., Ukkola-Vuoti, L., Karma, K., Onkamo, P., and Järvelä, I. (2016). Creative activities in music - a genome-wide linkage analysis. *PLoS One* 11, e0148679. <https://doi.org/10.1371/journal.pone.0148679>.
- Park, H., Lee, S., Kim, H.J., Ju, Y.S., Shin, J.Y., Hong, D., von Grotthuss, M., Lee, D.S., Park, C., Kim, J.H., et al. (2012). Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. *J. Med. Genet.* 49, 747–752. <https://doi.org/10.1136/jmedgenet-2012-101209>.
- Pfordresher, P.Q., and Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Percept* 25, 95–115. <https://doi.org/10.1525/MP.2007.25.2.95>.
- Pfordresher, P.Q., and Demorest, S.M. (2020). Construction and validation of the Seattle Singing Accuracy Protocol (SSAP): an automated online measure of singing accuracy. In *The Routledge Companion to Interdisciplinary Studies in Singing, Volume I: Development*, F.A. Russo, B. Ilari, and A.J. Cohen, eds. (Routledge). <https://doi.org/10.4324/9781315163734>.
- Pfordresher, P.Q., and Demorest, S.M. (2021). The prevalence and correlates of accurate singing. *J. Res. Music Educ.* 69, 5–23.
- Pfordresher, P.Q., Brown, S., Meier, K.M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am.* 128, 2182–2190. <https://doi.org/10.1121/1.3478782>.
- Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47, 702–709. <https://doi.org/10.1038/ng.3285>.
- Powell, S.J. (2014). Masculinity and choral singing: an Australian perspective. *Int. J. Music Educ.* 33, 233–243. <https://doi.org/10.1177/0255761414533307>.
- R Core Team (2020). *R: A Language Environment for Statistical Computing* (R Foundation for Statistical Computing).
- Revelle, W. (2019). *Psych: Procedures for Psychological, Psychometric, and Personality Research* (Northwestern University). <https://personality-project.org/r/psych/>.
- Seesjärvi, E., Särkämö, T., Vuoksima, E., Tervaniemi, M., Peretz, I., and Kaprio, J. (2016). The nature and nurture of melody: a twin study of musical pitch and rhythm perception. *Behav. Genet.* 46, 506–515. <https://doi.org/10.1007/s10519-015-9774-y>.
- Steele, C.J., Bailey, J.A., Zatorre, R.J., and Penhune, V.B. (2013). Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. *J. Neurosci.* 33, 1282–1290. <https://doi.org/10.1523/JNEUROSCI.3578-12.2013>.
- Tan, Y.T., Peretz, I., McPherson, G.E., and Wilson, S.J. (2021). Establishing the reliability and validity of web-based singing research. *Music Percept* 38, 386–405. <https://doi.org/10.1525/mp.2021.38.4.386>.

Thurstone, L.L. (1935). *The Vectors of Mind* (University of Chicago Press).

Ullén, F., Mosing, M.A., Holm, L., Eriksson, H., and Madison, G. (2014). Psychometric properties and heritability of a new online test for musicality, the Swedish Musical Discrimination Test. *Pers. Individ. Dif.* 63, 87–93. <https://doi.org/10.1016/j.paid.2014.01.057>.

Vaquero, L., Hartmann, K., Ripolles, P., Rojo, N., Sierpowska, J., Francois, C., Camara, E., van Vugt, F.T., Mohammadi, B., Samii, A., et al. (2016). Structural neuroplasticity in expert pianists depends on the age of musical training onset. *Neuroimage* 126, 106–119. <https://doi.org/10.1016/j.neuroimage.2015.11.008>.

Watts, C., Murphy, J., and Barnes-Burroughs, K. (2003). Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *J. Voice* 17, 185–194. [https://doi.org/10.1016/s0892-1997\(03\)00023-7](https://doi.org/10.1016/s0892-1997(03)00023-7).

Welch, G.F. (2015). Singing and vocal development. In *The Child as Musician: A Handbook of Musical Development*, 2nd, G.E. McPherson, ed. (Oxford University Press), pp. 441–461.

Wesseldijk, L.W., Mosing, M.A., and Ullén, F. (2019). Gene-environment interaction in expertise: the importance of childhood

environment for musical achievement. *Dev. Psychol.* 55, 1473–1479. <https://doi.org/10.1037/dev0000726>.

Wilson, S.J., Abbott, D.F., Lusher, D., Gentle, E.C., and Jackson, G.D. (2011). Finding your voice: a singing lesson from functional imaging. *Hum. Brain Mapp.* 32, 2115–2130. <https://doi.org/10.1002/hbm.21173>.

Wilson, S.J., Lusher, D., Martin, C.L., Rayner, G., and McLachlan, N. (2012). Intersecting factors lead to absolute pitch acquisition that is maintained in a “fixed do” environment. *Music Percept* 29, 285–296. <https://doi.org/10.1525/mp.2012.29.3.285>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
IBM SPSS Statistics (Version 27)	https://www.ibm.com/products/spss-statistics	RRID:SCR_016479
OpenMX (version 2.19.1)	https://openmx.ssri.psu.edu/ ; https://cran.r-project.org/web/packages/OpenMx/index.html	NA
TONY	https://code.soundsoftware.ac.uk/projects/tony	NA
R (version 4.0.4)	http://www.r-project.org/	RRID:SCR_001905
RStudio	https://www.rstudio.com/	RRID:SCR_000432
psych (version 2.1.3)	https://cran.r-project.org/web/packages/psych/index.html	RRID:SCR_021744
polycor	https://cran.r-project.org/web/packages/polycor/index.html	NA
Original code for analyses and figures	This paper	https://doi.org/10.26188/621c4f3d17610
Melbourne Singing Tool	Tan, Y. T., Peretz, I., McPherson, G. E., & Wilson, S. J. (2021). Establishing the reliability and validity of web-based singing research. <i>Music Perception</i> , 38(4), 386-405.	https://doi.org/10.1525/MP.2021.38.4.386

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Daniel Yeom (yeom.d@unimelb.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data that support the findings of this study are available from the [lead contact](#) upon reasonable request, and with the permission of Twins Research Australia. Any additional information required to re-analyze the data reported in this paper is available from the [lead contact](#) upon request.

All original code has been deposited at Figshare and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#). The standard version of TONY is freely available at <https://code.soundsoftware.ac.uk/projects/tony>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

1273 twins from across Australia were recruited for this study. 964 participants (75.7%) were recruited with assistance from Twins Research Australia (TRA), a nationwide twin registry that supports twin studies. Twins from the wider community were also recruited through social media, advertisements around the University of Melbourne and by word-of-mouth. The recruitment material specified that vocal ability was not a prerequisite to participate in the study to reduce the likelihood of self-selection from participants who may already have an interest in singing. As an incentive for participation, twin pairs that had both twins complete the entire study went into two successive prize draws; first for an iPad mini, and second for gift cards equivalent to AUD\$1000. This was done to reduce self-selection bias. Participants were informed that by taking part in the study, they were providing informed consent. Ethics was approved by the Human Research Ethics Committee of the University of Melbourne.

84 participants (6.6% of the original sample) were removed from the analyses due to missing values on at least one of the five singing measures. This meant they were not included in the calculation of the factor score, and therefore the heritability analyses. The final sample contained 1189 individuals (875 females). There were 377 MZ (279 female, 98 male) and 140 DZ (99 female, 30 male, 11 opposite-sex) twin pairs, as

well as 155 individual cases without their co-twin. The individual cases were retained for the univariate genetic analyses because they contribute to the estimation of means, variances and covariances.

The mean age of the final sample was 43.4 years (SD = 16.5) with a range of 15–90 years. On average, participants had 15.0 years of education (SD = 2.5, range = 8–18). 113 participants (9.5%) reported voice as their primary instrument, with a median of 5 years of training (1–3 years = 37.2%, 4–6 years = 28.3%, >6 years = 31.9%, missing = 2.7%). 755 (63.5%) reported a primary instrument other than voice, with a median of 5 years of training (1–3 years = 34.7%, 4–6 years = 30.1%, >6 years = 34.2%; missing = 1.1%). There was no significant difference in years of training between MZ and DZ twins ($F(1, 1121) = 0.727, p = 0.394$). A full breakdown of instrument categories in the sample is available in SI Appendix, [Table S1](#).

METHOD DETAILS

The Melbourne Singing Tool

Data for this study were collected using a purpose-built online singing tool, described in [Tan et al. \(2021\)](#). The tool is written in HTML5, ensuring compatibility with most modern smartphones, tablets, and computers, as well as most modern operating systems. [Participants](#) used their own devices to complete the study online and were asked to undertake the tasks in a quiet place to minimise background disruptions. Completion of the singing tool took 20 min on average and all singing recordings were sent to a secure server.

The Melbourne Singing Tool has been found to have high internal consistency, convergent validity, and test-retest reliability across an average 4.5 year period ([Tan et al., 2021](#)). In addition, similar performance of the singing tasks was shown by web-based [participants](#) to those who completed the tool in a controlled laboratory environment ([Tan et al., 2021](#)). In the current sample, 271 participants completed a pilot version of the singing tool, with 78 of these completing both the pilot and current versions, with singing performance highly correlated across the two versions (SI Appendix, [Table S2](#)).

The Melbourne Singing Tool first requires [participants](#) to provide informed consent, test their audio and recording set-up, and complete a demographic questionnaire. This is followed by two music perception tasks and the three main singing tasks used in this study, presented in the appropriate vocal range for the reported sex of the participant ([Table 1](#)). [Participants](#) then complete a series of self-report questionnaires about their early and current musical experiences and training. From the childhood and current environmental variables, three questions were chosen to reflect singing engagement most likely involving shared environmental experiences: (1) “I sang by myself [in childhood]” (1 = Never, 5 = Always), (2) “I sang with family [in childhood]” (1 = Never, 5 = Always) and (3) I sing with my family [in my current everyday life] (1 = Never, 5 = Always). To avoid practice effects, pilot data were used for the 78 participants who completed both versions of the task.

Singing variables

The singing data was processed using a modified version of the open-source *TONY* software ([Mauch et al., 2015; Tan et al., 2021](#)). *TONY* automatically identifies stable pitch segments and provides an initial estimate of each note’s median fundamental frequency, f_i ([Mauch et al., 2015](#)). Each participant’s singing data was then inspected both visually and aurally by a member of the research team using *TONY*’s graphical interface and manually adjusted to account for incorrectly identified notes, including octave errors, where necessary. Our previous work has shown that this method of supplementing *TONY*’s automatic segmentation with manual inspection correlates highly with purely manual pitch segmentation, thereby demonstrating its robustness and reliability ([Tan et al., 2021](#)). Five measures were extracted from the audio: Single Note PD, Familiar Tune PD and ID and Unfamiliar Tune PD and ID.

For the Single Note task, we calculated absolute PD as the absolute deviation from the target pitch in cents. To account for differences in duration of sung notes, we calculated PD using a 2-s segment from the onset of the note. If participants did not hold a note for 2 s, PD was calculated using the entire available segment. This was averaged over all trials, using the following equation:

$$PD = \frac{\sum_{i=1}^n \left| 1200 \times \log_2 \frac{f_i}{r_i} \right|}{n}$$

where f_i is the fundamental frequency (Hz) of the i th pitch produced by the participant, r_i is the fundamental frequency of the target pitch and n is the number of trials. For octave errors the following adjustment was used to prevent unfairly penalising participants:

$$PD_{\text{oct}} = \left| 1200 - 1200 \times \log_2 \frac{f_i}{r_i} \right|$$

where f_i is the fundamental frequency (Hz) of the i th pitch produced by the participant and PD_{oct} is the octave-adjusted PD. If TONY was unable to extract an estimate for f_i from a trial (either due to a blank audio recording or poor audio quality) this was treated as missing data, and subsequently excluded from calculations of the average PD for this task.

For the Unfamiliar Tune task, in addition to PD we calculated ID in terms of the absolute difference between each sung interval and its expected target interval in the tune using the following equation:

$$ID = \frac{\sum_{i=2}^n \left| \left(1200 \times \log_2 \frac{f_i}{f_{i-1}} \right) - \left(1200 \times \log_2 \frac{r_i}{r_{i-1}} \right) \right|}{n}$$

where f_i is the fundamental frequency (Hz) of the i th pitch produced by the participant, f_{i-1} is the fundamental frequency of the note preceding the i th pitch, r_i is the fundamental frequency of the target pitch, r_{i-1} is the fundamental frequency of the note preceding the target pitch and n is the number of notes in the task. Octave adjustments were applied as per the Single Note task. Where participants were missing notes audio files were inspected and the location of the missing note was determined from the contour of the sung tune. ID was then calculated between the i th pitch and the previous non-missing note.

For the Familiar Tune task, since participants chose their preferred starting note a reference pitch was first calculated for each trial before then applying the above process for calculating PD and ID. A stable reference pitch for each trial was estimated from the sung frequencies of the first, second and fourth notes of the first phrase of Happy Birthday (which share the same pitch), using the following equation:

$$r_i = 2^{\frac{int}{1200}} \times r_{i-1}$$

where r_i represents the fundamental reference frequency of the i th note in the tune, r_{i-1} is the fundamental reference frequency of the preceding note and int refers to the expected interval in cents between the two notes. Octave adjustments were not required as all expected pitch values were derived from the participant's starting note.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analyses of the raw singing data were conducted in R (version 4.0.4) (R Core Team, 2020) and the factor analysis was performed using SPSS version 27. Both the Kaiser-Meyer-Olkin test of sampling adequacy and Bartlett's test of sphericity suggested that the data was adequate for factor analysis (KMO = 0.79; $\chi^2 = 6841.56$, $p < 0.001$). To account for skew, each raw accuracy measure was first log-transformed. These values were then reverse-coded by multiplying by -1 and adding 7, with higher values representing higher accuracy on the singing tasks. Principal axis factoring was used as the extraction method, with no rotation since only one factor was extracted. The number of factors to retain was based on Kaiser's criterion (eigenvalues > 1) and inspection of the scree plot. Factor scores for the single latent factor were calculated using a Thurstone regression method (Thurstone, 1935).

All genetic analyses were conducted in R using the OpenMx (version 2.19.1) (Boker et al., 2011; Neale et al., 2016), psych (version 2.1.3) (Revelle, 2019) and polycor (version 0.7–10) (Fox and Dusa, 2019) packages. Assumption tests for quantitative and qualitative differences between sexes and zygosity were performed with an alpha criterion of $p = 0.01$. First, a saturated model was fitted to the data to estimate all parameters, allowing for means, variances and covariances to differ between first and second born twins, sexes and zygosity (i.e. Five groups – MZ female, MZ male, DZ female, DZ male and DZ opposite sex pairs). This full model was then compared against various constrained submodels. Means were first equated across twin order, and then by zygosity and sex. Variances were equated in the same order. Lastly, covariances were equated across sexes and then across zygosity. Model fitting statistics for all assumption

tests are shown in [Table 4](#). We then fitted a univariate ACE model with the singing factor score as the variable of interest, and age and sex as covariates ([Table 5](#)). Since assumption testing indicated that there were no sex differences, sex was not modeled separately.

To explore the potential contribution of shared family environment on singing ability, an exploratory multiple regression analysis was conducted in R, with age and sex included as covariates. Listwise deletion was used to remove participants with incomplete data on all predictors.