

# Genetic Correlates of Musical Pitch Recognition in Humans

Dennis Drayna,<sup>1\*</sup> Ani Manichaikul,<sup>1</sup> Marlies de Lange,<sup>2</sup>  
Harold Snieder,<sup>2†</sup> Tim Spector<sup>2</sup>

We used a twin study to investigate the genetic and environmental contributions to differences in musical pitch perception abilities in humans. We administered a Distorted Tunes Test (DTT), which requires subjects to judge whether simple popular melodies contain notes with incorrect pitch, to 136 monozygotic twin pairs and 148 dizygotic twin pairs. The correlation of DTT scores between twins was estimated at 0.67 for monozygotic pairs and 0.44 for dizygotic pairs. Genetic model-fitting techniques supported an additive genetic model, with heritability estimated at 0.71 to 0.80, depending on how subjects were categorized, and with no effect of shared environment. DTT scores were only weakly correlated with measures of peripheral hearing. This suggests that variation in musical pitch recognition is primarily due to highly heritable differences in auditory functions not tested by conventional audiologic methods.

The perception of pitch requires both the ear, which receives auditory signals, and the brain, which performs substantial processing of auditory signals to produce a perceived

pitch (1–3). Although the general features of human pitch processing have been well described, the precise cellular and molecular mechanisms involved remain largely obscure. One approach to understanding the mechanisms of pitch perception is to use genetic methods that exploit naturally occurring variation in pitch perception ability (4). If such variability is due to genetic factors, linkage and positional cloning studies could identify genes that encode the components of the pitch perception apparatus (5). To examine the genetic contributions to musical pitch

---

<sup>1</sup>National Institute on Deafness and Other Communication Disorders, National Institutes of Health, 5 Research Court, Rockville, MD 20850, USA. <sup>2</sup>Twin Research and Genetic Epidemiology Unit, St. Thomas' Hospital, London, SE1 7EH, UK.

\*To whom correspondence should be addressed.

†Present address: Georgia Prevention Institute, Medical College of Georgia, Building HS-1640, Augusta, GA 30912, USA.

recognition ability in humans, we performed a twin study (6) using the Distorted Tunes Test (DTT), which requires subjects to recognize notes with incorrect pitch in simple popular melodies (7).

The original DTT was developed in the 1940s (8) and used in large studies in the British population. These studies suggested that cultural biases and the effects of musical experience could be minimized by the appropriate choice of melodies, and that test scores in the same individual were stable across decades. They also revealed that a small but significant portion of the population (about 5%) scored no better than chance in their ability to distinguish correct from incorrect melodies. These individuals were classified as "tune deaf."

For our study, we created a similar, updated DTT and validated it for use in the current U.S. and British populations (9). The updated DTT was recorded on a compact disk and presented to all subjects in the same setting. Briefly, subjects were presented with 26 short popular melodies, ranging in length from 12 to 26 notes. Tunes were presented once, and after each presentation, subjects were asked to score whether the melody was correct or incorrect, and whether they were familiar or unfamiliar with that melody. We first measured the performance of 50 unrelated males and 50 unrelated females on the updated DTT. The distribution of scores in males and females did not differ (Kolmogorov-Smirnov = 0.94, Mann-Whitney test = 0.78). Test-retest scores in the same subject were highly correlated ( $n = 40$ ,  $r = 0.77$ ), confirming that like the original DTT, the updated DTT is reproducible in individuals. In contrast to results obtained by Kalmus and Fry with the original DTT, we did not observe a clear distinction between tune deaf and normal individuals.

Because our goal was to determine the

extent to which genes and/or environment influence musical pitch-recognition ability, we chose a twin study, which can discriminate between the effect of the shared environment and that of shared genes. The study was approved by the St. Thomas' Hospital Research Ethics Committee (EC95/041, modification approved 29 September 1999), and informed consent was obtained from all subjects. A total of 284 female Caucasian twin pairs [136 identical (monozygotic, MZ) and 148 nonidentical (dizygotic, DZ)] aged 18 to 74 years from the St. Thomas' UK Adult Twin Registry (10) participated in the study. Subjects in this registry were ascertained from the general population through national media campaigns in the United Kingdom. Participating twins were part of an ongoing study into the genetics of common complex diseases (10–13). Twins were unaware of the specific hypotheses tested and were not screened for IQ, musical training, or musical experience.

The median ages of the MZ group and the DZ group were 50.7 and 47.9 years, respectively. Zygosity was determined by standardized questionnaire, with DNA fingerprinting used for confirmation (10). Each subject was administered the DTT and the 5 Minute Hearing Test (FMHT) to help identify subjects with potentially confounding hearing loss. The FMHT, promulgated by the American Academy of Otolaryngology, has been widely used for initial screening for hearing loss, and high correlations have been reported with a wide range of hearing measures, including pure-tone audiometry (14).

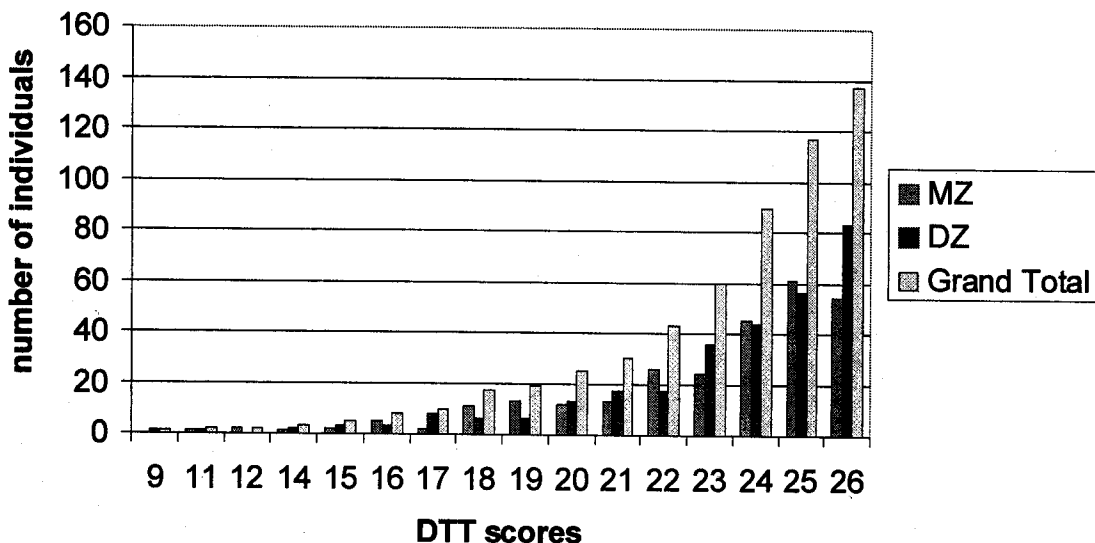
Using the DTT, we measured musical pitch recognition ability on an ordinal scale, scored as the number of correctly classified tunes. The scores of the subjects ranged from 26 (a perfect score) to 9. The distribution of scores in the MZ and DZ pairs is shown in Fig. 1, along with the overall score distribu-

tion of the entire sample. Although there was a trend for the MZ twins to have lower scores, this difference was not statistically significant (chi-square,  $P = 0.12$ ; Kolmogorov-Smirnov,  $P = 0.26$ ).

The statistical package STATA (15) was used to analyze the data. Spearman correlations were used to describe the associations between the scores on the DTT and the FMHT. A slight negative correlation between the DTT and the FMHT was observed (Spearman's  $r = -0.10$ ,  $P = 0.01$ ). Despite having statistical significance, this small negative correlation may or may not have functional significance. To test if this affected scores on the DTT, we divided subjects into a hearing and a hearing-impaired group on the basis of their FMHT scores (hearing impaired = FMHT score >15). No difference in DTT score was found between the groups as determined with statistical methods that account for the nonindependence of the twin pairs (generalized estimating equation,  $P = 0.45$ ). There was no correlation between the DTT and age (Spearman's  $r = -0.01$ ,  $P = 0.87$ ). Probandwise concordance rates for the data divided into the two original categories described by Kalmus and Fry were 0.75 for MZ and 0.57 for DZ twin pairs, and a test for the difference was significant ( $P = 0.01$ ) (16), indicating a genetic influence.

We applied genetic model-fitting techniques using the structural equation modeling package Mx (17) to obtain estimates of the genetic and environmental factors. Model fitting allows separation of the observed phenotypic variance into additive (A) or dominant (D) genetic components and common (C) and unique (E) environment. E also contains measurement error. The heritability, which estimates the extent to which variation in liability to disease in a population can be explained by genetic variation, can be defined as the ratio of genetic variance (A + D)

**Fig. 1.** Distribution of twin scores on the Distorted Tunes Test. Bars indicate the number of subjects attaining each score on the DTT. MZ, scores of monozygotic twins; DZ, scores of dizygotic twins.



to total phenotypic variance ( $A + C + D + E$ ). We tested the significance of A, C, and D by removing them sequentially in specific submodels and testing the deterioration in model fit after each component was dropped from the full model. This leads to a model explaining the data with as few parameters as possible. Standard hierarchical chi-squared tests were used to select the best-fitting model (18).

The genetic and environmental contributions for categorical traits can be quantified by assuming a continuous underlying normal liability distribution with multiple thresholds that discriminate between the categories (18). We estimated the correlation in liability within the twin pairs by deriving polychoric correlations from the pairwise categorical distribution. The model-fitting approach compares the size of the polychoric correlations in MZ and DZ twins and provides estimates of the relative contribution of genetic and environmental factors to the liability distribution underlying musical pitch perception ability (18, 20). Using the original definition of Kalmus and Fry (tune deaf = a score  $\leq 23$ ) on our data resulted in a large percentage of tune deaf individuals (39.6%). Using the updated DTT, we also did not observe a clear distinction between tune deaf and normal individuals. We therefore also fitted the data without imposing any arbitrary cut-off points and assigned categories as the raw scores themselves, with the exception of the few lowest scorers (scores between 9 and 15) that were assigned to one category, resulting in 12 categories. Contingency tables for the 12 categories were produced for the MZ and DZ twins and used as input data for Mx.

Analysis of the complete data (with no arbitrary cut-off point) showed a correlation in liability within the twin pairs of 0.67 and 0.44 for MZ and DZ twins, respectively (see Table 1, Model 1 for twin correlations in liability on the basis of two category data).

Across the analyses, the model providing the best fit to the data included an additive genetic and a unique environmental component. The heritability as estimated by genetic model-fitting with all of the available data was 71% [95% confidence interval (CI): 61 to 78%]. Using the original cut-off value of

**Table 1.** Twin correlations and heritabilities of best-fitting models for the DTT data with two categories (Model I) or the actual scores (Model II). *r*MZ, monozygotic twin correlation; *r*DZ, dizygotic twin correlation; *h*<sup>2</sup>, heritability; 95% CI, 95% confidence interval.

Model	Thresholds	<i>r</i> MZ	<i>r</i> DZ	<i>h</i> <sup>2</sup> (95% CI)
I	$\leq 23$	0.79	0.46	80% (65–90%)
II	$\leq 15$ , actual score	0.67	0.44	71% (61–78%)

Kalmus and Fry ( $\leq 23$ ) to define two classes corresponds to a simplified model that contains only two groups: those with normal pitch recognition and those with some deficit in pitch recognition, regardless of severity. Using this model, we estimated heritability at 80% (95% CI: 65 to 90%). In both analyses, no dominant genetic effects and no significant effect of shared environment were detected (Table 1).

Despite the major role of genetic factors underlying DTT scores, a certain amount of musical experience is nevertheless required to perform well on the DTT, and the original results of Kalmus and Fry provided evidence for the effects of such experience (7, 19). Any effects of culture or musical experience are likely to be the same for both MZ and DZ twins and thus should have no effect on the heritability estimates that we have found. Our data indicate that individual differences in musical experience may be at least partly responsible for the fraction of variance in DTT scores attributable to unique environment (E), 20 to 29%. Because of the DTT's requirement for musical experience, it may be a conservative measure of the heritability of variation in pitch perception in isolation, that is, measured in a way that does not require such experience.

Because the FMHT serves as a rough measure of peripheral hearing, its poor correlation with the DTT suggests that musical pitch perception is largely independent of peripheral hearing and that variation in pitch perception originates in portions of the auditory system that are independent of this function.

Melodies consist of a series of tones presented in a specific order and rhythm, in which successive tones differ by specific intervals. In the DTT, the note order and rhythm remain unchanged, and only the interval between successive tones is altered. Because variation in long-term tonal memory (9) and musical experience appear to have modest effects, the DTT primarily determines a subject's ability to measure successive pitch intervals. Although the DTT has a number of characteristics that make it ideal for a large-scale twin study, it will be important to use other measures of pitch recognition to confirm the results obtained with the DTT. Similarly, indistinguishable distributions of DTT scores in males and females suggest that results from female subjects can be generalized to the whole population, but it will be important to confirm this with male subjects.

The original DTT was designed to identify individuals with severe deficits in pitch recognition. Our results with the modified DTT indicate a high heritability for performance across the full spectrum of pitch recognition abilities in the general population. Studies elsewhere have demonstrated a sig-

nificant genetic contribution to absolute pitch (AP), a relatively rare phenomenon in which individuals are capable of identifying a particular tone without the use of a reference tone. However, AP individuals do not necessarily have superior pitch acuity compared to those without AP (20). In addition, the full development of AP apparently has a strong environmental component, with specific musical training (perhaps during a critical developmental time period) being required for expression (21). Thus, AP stands in contrast to the ability measured by the DTT, and it is not clear if the genetic factors in AP have any relation to those that underlie DTT scores.

The heritability estimates we observe for this measure of deficits in pitch recognition are very substantial and are as high as or higher than those for many common complex traits in humans (22). However, these high heritability estimates do not address several important issues, including the number of genes involved and their relative effects. Our results indicate that genetic approaches, which are ideal for studying traits with unknown biochemical or cellular mechanisms, are likely to be fruitful in efforts to understand this neural function.

#### References and Notes

- B. C. J. Moore, in *Springer Handbook of Auditory Research: Human Psychophysics*, W. A. Yost, A. N. Popper, R. R. Fay, Eds. (Springer-Verlag, New York, 1993), pp. 56–115.
- A. J. M. Houtsuma, in *Handbook of Perception and Cognition: Hearing*, B. C. J. Moore, Ed. (Academic Press, New York, 1996), pp. 267–295.
- W. M. Hartmann, *Signals, Sound and Sensation* (American Institute of Physics, Woodbury, NY, 1997), pp. 117–145.
- D. Drayna, *Nature Genet.* **18**, 96 (1998).
- F. Collins, *Nature Genet.* **9**, 347 (1995).
- T. D. Spector, H. Snieder, A. J. MacGregor, *Advances in Twin and Sib-Pair Analysis* (Greenwich Medical Media, London, 2000).
- H. Kalmus, D. Fry, *Ann. Hum. Genet.* **43**, 369 (1980).
- D. B. Fry, *Speech* (March 1948), p. 1.
- The original DTT consisted of 26 short melodies, ranging in length from 12 to 26 notes. Ten of these melodies were played normally, whereas 16 contained tonal errors (7). Our updated test contained 8 of the original DTT melodies verbatim, plus 18 new melodies, for a total of 26, which ranged from 12 to 26 notes in length. Of these, 9 were played correctly, whereas 17 were distorted to produce errors. The errors were introduced according to the general rules used by Kalmus and Fry in their original DTT, changing the pitch of two to nine notes, generally within one or two semitones of the correct note and following the overall rise and fall of the normal melody. As in the original DTT, all melodies were unaltered in rhythm and note order. A list of the melodies and the errors introduced is available at [www.nidcd.nih.gov/intram/scientists/draynad.htm](http://www.nidcd.nih.gov/intram/scientists/draynad.htm). The melodies were produced in pure tones with Mozart version 3.2 software (D. Webber, Mozart Music Software, Warrington, UK) on a Macintosh G3, rendered more natural by the addition of overtones with Arnold's MIDI (Musical Instrumentation Digital Interface) player software. Ultra Recorder version 2.4 (E. J. Campbell, EJ Enterprises, <http://members.aol.com/EJC3/>) was used to convert the MIDI files to AIFF (Audio Interchange File Format) and, after joining voice-over spoken instructions with SndSampler, was recorded along with spoken instructions on a

compact disk with Adaptec Toast 3.5.4. Copies of the updated DTT on a compact disk are available from the corresponding author. The updated DTT contained a preponderance of tunes played incorrectly (17/26), but our data suggest that whether a tune was played correctly or incorrectly did not significantly bias subjects' answers. Four tunes were played twice in the updated DTT; each was played correctly once and incorrectly once. Wrong answers on these eight questions were evenly distributed; 51% were correct melodies identified as incorrect, and 49% were incorrect melodies identified as correct. Two methods were used to determine the role of long-term memory and cultural experience in the performance on the updated DTT. First, subjects were asked whether they were familiar with each tune presented, and familiarity was positively correlated with correct answers. We also developed a second test, termed the International Tunes Test, that consisted of 18 short melodies chosen to be unfamiliar to all subjects and that used both Western and non-Western tonal systems. Subjects were presented with each melody played correctly twice in succession and were then asked whether a third rendition played immediately thereafter was the same as or different than the first two playings. Six of the International Tunes were played correctly the third time, whereas 12 were played incorrectly the third time. The distribution of scores on the International Test were indistinguishable from those on the updated DTT, with no significant differences between males and females. Individuals' scores on the DTT and the International Test were highly correlated ( $r = 0.71$ ). These results suggest that performance on the updated DTT is not highly dependent on long-term musical memory.

10. T. Spector *et al.*, *Br. Med. J.* **312**, 940 (1996); C. J. Hammond *et al.*, *N. Engl. J. Med.* **342**, 1786 (2000).
11. H. Snieder *et al.*, *Hypertension* **35**, 574 (2000).
12. V. Bataille, H. Snieder, A. J. MacGregor, P. Sasieni, T. D. Spector. *J. Natl. Cancer Inst.* **92**, 457 (2000).
13. M. E. de Lange, H. Snieder, R. A. S Ariëns, T. D. Spector, P. J. Grant. *Lancet* **357**, 101 (2001).
14. K. Koike *et al.*, *Otolaryngol. Head Neck Surg.* **111**, 625 (1994).
15. StataCorp 1997 Stata Statistical Software, Release 5.0 (Stata Corporation, College Station, TX).
16. J. S. Witte, J. B. Carlin, J. L. Hopper, *Genet. Epidemiol.* **16**, 290 (1999).
17. Mx, version 4; M. C. Neale, Medical College of Virginia, Virginia Commonwealth University.
18. M. C. Neale, L. Cardon, *Methodology for Genetic Studies of Twins and Families* (Kluwer, Dordrecht, Netherlands, 1992). The chi-square for the best-fitting model (additive genetic component and unique environmental component same thresholds) = 256.810 (df = 274), with  $P = 0.765$ , indicating an excellent fit of the model to the data.
19. D. Levitin. *Percept. Psychophys.* **56**, 414 (1994).
20. J. Profita, G. Bidder, *Am. J. Med. Genet.* **29**, 763 (1988); A. H. Takeuchi, S. H. Hulse, *Psychol. Bull.* **113**, 345 (1993).
21. S. Baharloo, P. A. Johnston, S. K. Service, J. Gitschier, N. B. Freimer, *Am. J. Hum. Genet.* **62**, 224 (1998); P. K. Gergesen *et al.*, *Am. J. Hum. Genet.* **65**, 911 (1999).
22. A. J. MacGregor, H. Snieder, N. J. Schork, T. D. Spector, *Trends Genet.* **16**, 131 (2000).
23. We thank the research nurses for skillful data collection and especially the twin volunteers who participated in this study. We also thank E. Balaban and the anonymous reviewers for their insightful comments. Supported by NIH grant Z01-DC-00043-03 from the National Institute on Deafness and Other Communication Disorders (D.D. and A.M.). M.d.L. and H.S. are sponsored by the British Heart Foundation (grants FS/99010 and FS/99050). The Twins Research Unit gratefully acknowledges support from the Arthritis and Rheumatism Council, Wellcome Trust, British Heart Foundation, and Gemini Genomics.