Please cite this article in press as: Roswandowitz et al., Two Cases of Selective Developmental Voice-Recognition Impairments, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.08.048

Current Biology 24, 1–6, October 6, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.08.048

Report

Two Cases of Selective Developmental Voice-Recognition Impairments

Claudia Roswandowitz,[1,2,*] Samuel R. Mathias,[3] Florian Hintz,[4,5] Jens Kreitewolf,[1] Stefanie Schelinski,[1] and Katharina von Kriegstein[1,6]

[1]Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

[2]International Max Planck Research School on Neuroscience of Communication, 04103 Leipzig, Germany

[3]Department of Psychiatry, Yale University, New Haven, CT 06511, USA

[4]Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, the Netherlands

[5]International Max Planck Research School for Language Sciences, 6500 AH Nijmegen, the Netherlands

[6]Department of Psychology, Humboldt Universität zu Berlin, 12489 Berlin, Germany

Summary

Recognizing other individuals is an essential skill in humans and in other species [1–3]. Over the last decade, it has become increasingly clear that person-identity recognition abilities are highly variable. Roughly 2% of the population has developmental prosopagnosia, a congenital deficit in recognizing others by their faces [4]. It is currently unclear whether developmental phonagnosia, a deficit in recognizing others by their voices [5], is equally prevalent, or even whether it actually exists. Here, we aimed to identify cases of developmental phonagnosia. We collected more than 1,000 data sets from self-selected German individuals by using a web-based screening test that was designed to assess their voice-recognition abilities. We then examined potentially phonagnosic individuals by using a comprehensive laboratory test battery. We found two novel cases of phonagnosia: AS, a 32-year-old female, and SP, a 32-year-old male; both are otherwise healthy academics, have normal hearing, and show no pathological abnormalities in brain structure. The two cases have comparable patterns of impairments: both performed at least 2 SDs below the level of matched controls on tests that required learning new voices, judging the familiarity of famous voices, and discriminating pitch differences between voices. In both cases, only voice-identity processing per se was affected: face recognition, speech intelligibility, emotion recognition, and musical ability were all comparable to controls. The findings confirm the existence of developmental phonagnosia as a modality-specific impairment and allow a first rough prevalence estimate.

Results and Discussion

"Phonagnosia" refers to a selective deficit in voice-identity recognition, which is dissociable from other forms of person recognition (e.g., via faces or names) and other aspects of voice perception (e.g., emotion or speech processing) [6–9]. Recently, the first putative case of developmental phonagnosia was reported [5]. However, in addition to performing poorly on several voice-recognition tasks, she also had difficulties with understanding speech in noise compared to controls. Therefore, it is currently still unclear whether developmental phonagnosia actually exists as a specific disorder that is dissociable from other complex auditory abilities. Confirming the existence of phonagnosia would have important implications for long-standing models of person perception [7]. A central assumption of these models is that voice recognition dissociates from our ability to understand what is said (speech recognition). However, this dissociation has recently been called into question (e.g., [10]). Thus, finding phonagnosia cases in which speech recognition is intact would advance our understanding of fundamental mechanisms in person recognition.

We employed a four-stage screening procedure to identify cases of developmental phonagnosia (Figure 1A; for a complete description of the screening procedure and results, see Supplemental Experimental Procedures available online). First, we developed a web-based test (http://www.phonagnosie.de) to assess participants' abilities to learn and recognize new voices (Figure 1B, "voice-name test"); this resulted in 1,057 complete data sets (Table S1). We sent 233 participants, who either (1) performed at least 1.5 SDs below the laboratory control mean or (2) rated themselves as poor voice recognizers, a detailed follow-up questionnaire. Of the 55 responses we received, the responses of ten individuals seemed to be indicative of a selective deficit in voice recognition. After a semistructured telephone interview, four individuals were invited to laboratory testing. They completed a behavioral test battery, an audiometric hearing test, a general neuropsychological assessment, and a structural MRI scan. Two of these participants (AS and SP) had no history of neurological or psychiatric illness, had normal hearing, performed within the normal range on the neuropsychological assessment (Table S2), and showed no pathological abnormalities in their MR images. For both AS and SP, we invited separate control groups who were matched in gender, age, education, and handedness [11] (see Supplemental Experimental Procedures). We compared differences between the scores of AS and SP and their respective control groups by using a modified t test [12], a standard procedure for comparing single cases to control groups (see, e.g., [5, 9, 13]). Differences with a probability $p < 0.05$ were considered statistically significant. The study was approved by the Research Ethic Committee of Leipzig University.

AS and SP Were Impaired in Voice Recognition

AS and SP are both 32 years old, highly educated, and, apart from their voice-recognition difficulties, completely healthy. Their lifelong voice-recognition difficulties include recognizing celebrities, close friends, and family members. For example, AS, a graduate student, finds it difficult to recognize her own daughter's voice when she is playing with her friend in another room. SP, a PhD student, first became aware of his difficulties when watching television with a friend: unlike his friend, SP failed to notice when a voice actor from his favorite dubbed television series changed. When answering telephone calls, SP often relies on compensatory strategies to

*Correspondence: roswandowitz@cbs.mpg.de

Please cite this article in press as: Roswandowitz et al., Two Cases of Selective Developmental Voice-Recognition Impairments, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.08.048

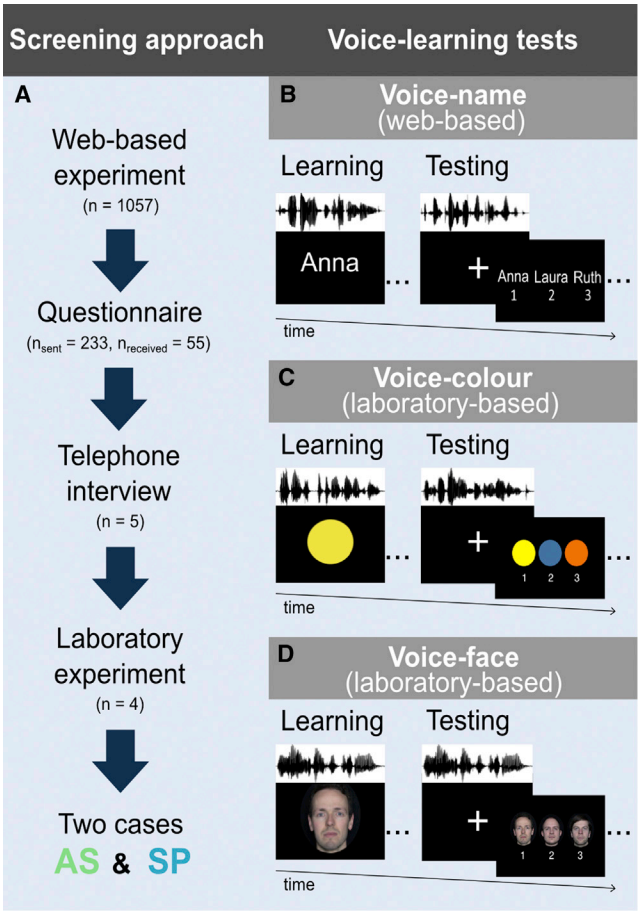


Figure 1. Screening Approach and General Procedure of the Voice-Learning Tests

(A) We assessed voice-recognition abilities in a wide audience using a web-based voice-name test published on http://www.phonagnosie.de, a paper-based questionnaire, semistructured telephone interviews, and laboratory-based testing. We identified AS and SP as two cases with developmental phonagnosia.

(B) Voice-name test. Participants learned six unfamiliar voices (three female and three male, displayed here as amplitude waveform) in association with a first name. After learning, participants were tested on their voice-recognition abilities. In each trial, they listened to a previously learned voice and selected the speaker's name among three alternatives presented on a screen via button press.

(C) Voice-color test. This test was structured similarly to the voice-name test, except that the names were replaced by colors.

(D) Voice-face test. This test was structured as the other two voice tests, except that now faces were learned together with the voices.

All three tests were based on different speakers' voices. The voice samples always differed between the learning and testing phases, i.e., learning and testing were done with different sentence material.

guess who is speaking, including making use of context and speaking characteristics (e.g., pauses, speaking rate). Both AS and SP feel embarrassed when they fail to identify familiar voices.

Besides the web-based voice-name test (Figure 1B), AS and SP performed two additional voice-learning tests in the laboratory: in the "voice-color test" (Figure 1C), participants learned to associate six unfamiliar voices with colors, and in the "voice-face test" (Figure 1D), they learned to associate the six unfamiliar voices with photographic images of their faces. Each test included a different set of speakers so that none of the speakers was heard in more than one test. AS and SP performed at least 2 SDs worse than controls on two of the three voice-learning tests and at least 1.5 SDs worse on the other test (Table S3). AS scored 50% on the voice-name test (controls: 75%) and 47% on the voice-color test (controls: 74%); these differences were statistically significant (voice-name test: $p = 0.016$; voice-color test: $p = 0.006$). AS also performed poorly on the voice-face test. This difference was close to statistical significance (AS: 73%; controls: 87%; $p = 0.067$) (Figure 2A; Table S3). SP scored 55% on the voice-name test (controls: 80%), 47% on the voice-color test (controls: 77%), and 67% on the voice-face test (controls: 90%). On two tests, the differences between SP and controls were significant (voice-color: $p < 0.001$; voice-face: $p < 0.001$), and the difference on the voice-name test was close to significance ($p = 0.064$) (Figure 2A; Table S3).

We also tested how well AS and SP recognized famous voices. In the "famous voice-recognition test," participants categorized the familiarity of voice samples of familiar and unfamiliar people and additionally provided the names of those categorized as familiar (see Supplemental Experimental Procedures). Both AS and SP showed a significant bias (c; [14]) in their familiarity judgments, compared to controls (AS: $p = 0.001$; SP: $p = 0.01$). For AS, accuracy (d′) was also significantly poorer ($p = 0.002$) (Figure 2B; Table S3). Although AS had problems with classifying voices as familiar, she nevertheless performed well at naming the ones she did classify correctly ($p = 0.33$). Conversely, SP performed worse at naming those he accurately classified as familiar, relative to his controls ($p < 0.001$) (Table S3).

AS and SP Performed Normally on Auditory and Visual Control Tests

To test whether their voice-recognition deficits were selective, AS and SP performed several control tests (see Supplemental Experimental Procedures). In the "speech-in-noise test," participants listened to subject-verb sentences mixed with noise and selected which verb was spoken out of four alternatives. AS's and SP's speech-reception thresholds were similar to controls (Figure 2C; Table S4), indicating that they were able to understand speech in noise normally. In the "vocal-emotion test," participants listened to words spoken in different affective states and selected the target emotion from six alternatives. AS and SP performed normally on this test as well (Figure 2C; Table S4). AS and SP also performed normally on two tests of face recognition, i.e., the Cambridge Face Memory Test [15] and a novel "face-name learning test" (Figure 2C; Table S4). A formal comparison of the performance on the voice and control tests revealed that AS and SP have a selective deficit in voice recognition (Figure 2D).

Apperceptive or Associative Phonagnosia?

Agnosias are classically divided into two forms: an apperceptive form and an associative form (for review, see [16]). "Apperceptive" agnosia refers to a failure to integrate the physical characteristics of a stimulus into a coherent percept, whereas "associative" agnosia refers to a failure to associate semantic information with the stimulus even when the stimulus itself is perceived normally. To disentangle the two forms of agnosia, we used a "voice-discrimination test" (see Supplemental Experimental Procedures). Participants listened to pairs of sentences spoken by three unfamiliar voices. After each pair, participants decided whether the sentences were spoken by the same speaker or not. AS performed 8.73 SDs below

Please cite this article in press as: Roswandowitz et al., Two Cases of Selective Developmental Voice-Recognition Impairments, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.08.048

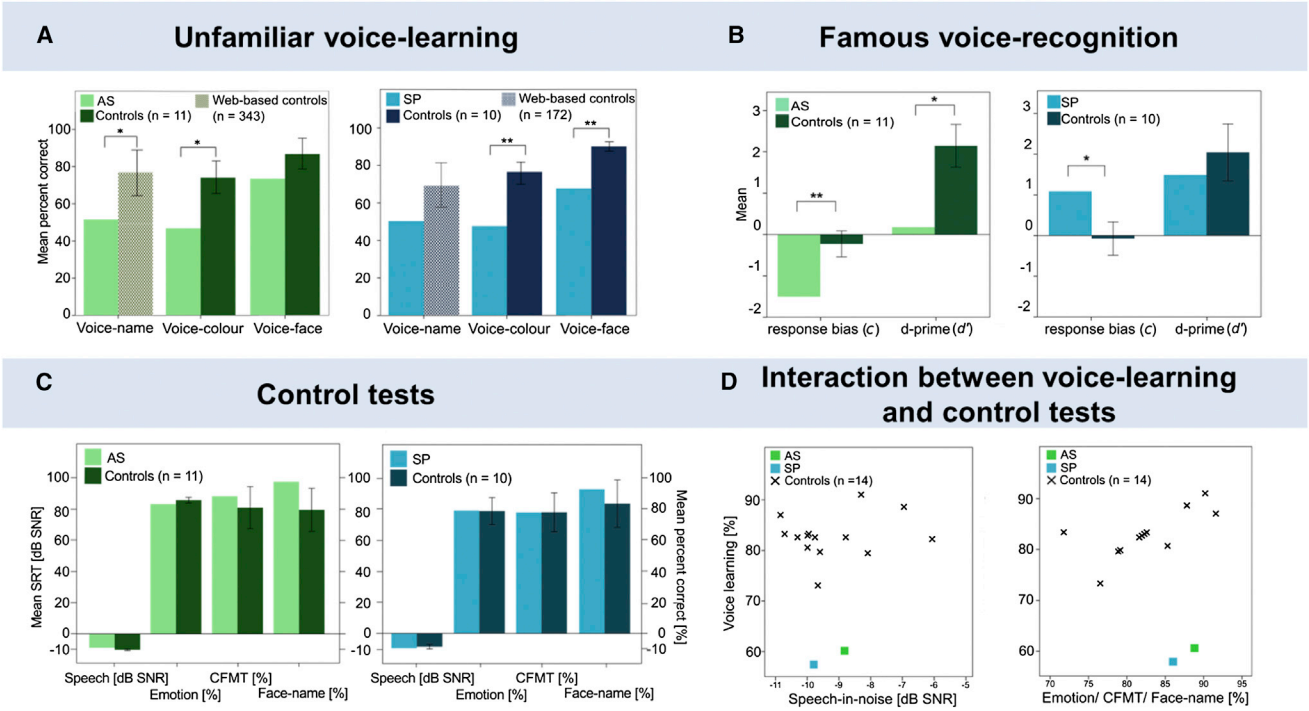


Figure 2. Results of AS and SP and Their Control Groups in the Voice and Control Tests

(A) Performance on the three voice-learning tests.

(B) Performance on the famous voice-recognition test. We computed response bias (c) on familiarity decision and indices of sensitivity (d′) to classify voices into famous and nonfamous categories.

(C) Speech receptive thresholds (SRT) derived from signal-to-noise ratios (dB SNR) on the speech-in-noise test and mean percent corrects on the overall performance of the vocal-emotion test, the Cambridge Face Memory Test (CFMT), and the face-name test are displayed. There were no differences between AS/SP and their respective control groups.

(D) The plots show AS's and SP's scores on the voice-learning tests (mean performance of laboratory voice-color and voice-face tests) in relation to their performance on the control tests. The interaction between the voice-learning and speech-in-noise performances is plotted in the left panel, and the interaction between the voice-learning and emotion/CFMT/face-name (mean performance over all three tests) performances is plotted in the right panel. To test the task type (voice, speech) and group (AS/SP, controls) interaction, we computed the mean differences between the voice-learning and the speech-in-noise performances separately for AS, SP, and controls. The differences were statistically significant (voice, speech: AS: $p < 0.001$, SP: $p < 0.001$; voice, emotion/CFMT/face-name: AS: $p = 0.002$, SP: $p = 0.002$). The control group included controls (of AS and SP) who completed all laboratory tests (see Supplemental Experimental Procedures).

All error bars show 1 SD. Asterisks indicate statistical significant differences among AS, SP, and their respective control groups (voice-name test: web-based controls; all remaining tests: laboratory-based controls) (*$p < 0.05$, **$p < 0.001$). See also Tables S3 and S4.

her controls ($p < 0.001$) (Figure 3A; Table S3). This suggests that she has apperceptive phonagnosia because this test does not require association with semantic information. By contrast, SP's performance was within the normal range ($p = 0.47$) (Figure 3A; Table S3). Thus, SP seems to have an associative phonagnosia. These patterns of impairment are consistent with AS's and SP's performances on the famous voice-recognition test, where only SP failed to name the voices he correctly categorized as familiar.

AS and SP Were Impaired in Vocal Pitch, but Not Vocal Timbre Perception

Pitch and timbre are two basic acoustic properties that provide important information for both discriminating unfamiliar voices [18] and recognizing familiar ones [19]. We therefore measured AS's and SP's just-noticeable differences (JNDs) for pitch and timbre (see Supplemental Experimental Procedures). In the "vocal-pitch test," participants listened to pairs of vowels and reported which had the higher pitch. The stimuli were resynthesized versions of the same original vowel [20], differing only in their fundamental frequency. In the "vocal-timbre test," participants listened to pairs of vowels and reported which was spoken by the smaller speaker. The stimuli were resynthesized versions of the same vowel [20], differing only in their "vocal-tract length," an aspect of vocal timbre that provides information about speaker size [21]. We found a clear dissociation between pitch and timbre JNDs (Figures 3B and 3C; Table S5). AS's and SP's pitch JNDs were around 3 SDs larger than controls' pitch JNDs, and they were well over one semitone, indicating severe impairments in pitch perception (AS: $p = 0.014$; SP: $p = 0.004$). However, their timbre JNDs indicated normal timbre perception. Thus, the impairments in voice-identity recognition observed in AS and SP coincided with severe deficits in pitch perception. It is highly unlikely that these deficits were caused by abnormalities at the level of the cochlea: AS performed normally, and SP performed better than controls on a test of cochlear function ("notched-noise test"; Figure S1, Table S5, and Supplemental Experimental Procedures). We also found that both AS and SP had normal JNDs for amplitude-modulation rate discrimination (for subtest-specific performance, see Figure S1, Table S5, Supplemental Experimental Procedures, and Supplemental Discussion).

Please cite this article in press as: Roswandowitz et al., Two Cases of Selective Developmental Voice-Recognition Impairments, Current Biology (2014), http://dx.doi.org/10.1016/j.cub.2014.08.048

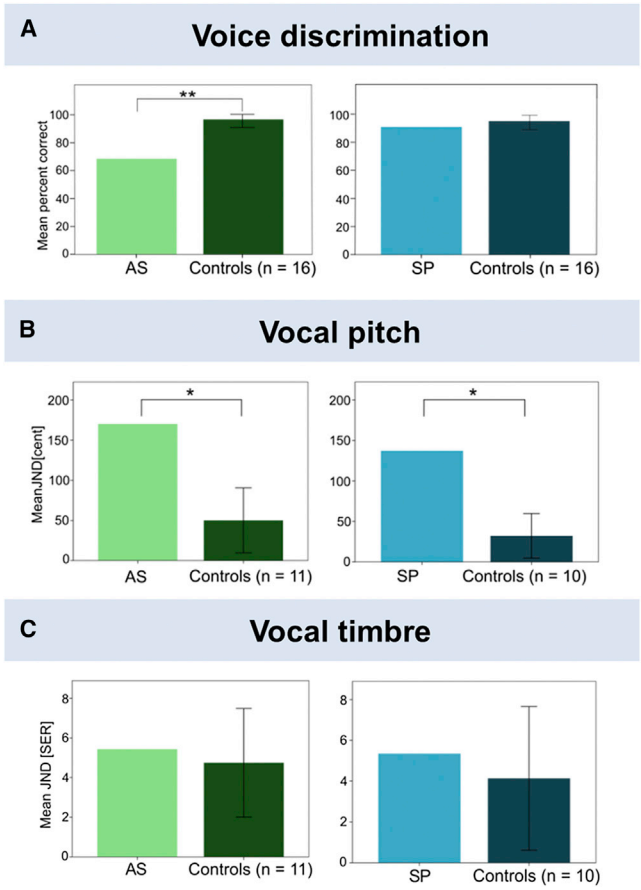


Figure 3. Performance on the Voice-Discrimination Test and Tests Assessing Pitch and Timbre Processing

(A) Performance on the voice-discrimination test. See also Table S3.

(B) Vocal-pitch test. Cent describes the logarithmic unit for F0 (fundamental frequency) intervals. See also Table S5.

(C) Vocal-timbre test. There were no differences between AS/SP and their respective control groups. Spatial envelope is a unit for the acoustic effect of the speaker's vocal-tract length [17]. SER, spatial envelope ratio. See also Table S5.

All error bars show 1 SD. Asterisks indicate statistical significant differences among AS, SP, and their respective control groups (*$p < 0.05$, **$p < 0.001$). JND, just-noticeable difference.

AS and SP Are Not Amusic

In addition to its role in voice-identity recognition [18], pitch perception is integral to the perception of music [22], and pitch JNDs greater than one semitone can be a symptom of congenital amusia, a lifelong disability in perceiving music [23]. Poor pitch perception is the most common deficit found in congenital amusics [23, 24], although they can be impaired in timbre perception as well [25]. Several lines of evidence suggest that neither AS nor SP is amusic. First, during a structured interview (Table S6), both reported being good at detecting when someone else sings out of tune and recognizing a familiar melody without the help of lyrics—two skills that are indicative of amusia [26]. Second, AS and SP performed within the normal range on an online version [26] of the Montreal Battery for the Evaluation of Amusia [27], the standard tool for diagnosing amusia (Figure 4A; see Table S7 for performance on specific subtests). Third, AS and SP performed normally on a test of musical instrument recognition (Figure 4B; see Supplemental Experimental Procedures).

Poor pitch discrimination is not always symptomatic of amusia; some completely healthy individuals can have abnormally large pitch JNDs but otherwise normal hearing and music appreciation [28]. Whether these individuals suffer from voice-recognition impairments is currently unknown. Prosopagnosics are often unaware of their face-recognition deficit [29], and this might also be the case in phonagnosia. In congenital amusia, pitch-perception deficits are often assumed to be the primary cause of the disturbances in perceiving music [30]. If developmental phonagnosia is consistently associated with a pitch discrimination deficit, it therefore might be of a different nature than the one associated with amusia. Alternatively, a pitch discrimination deficit per se might not cause phonagnosia, but it might nevertheless exacerbate poor voice-recognition abilities.

Prevalence of Developmental Phonagnosia

Because we collected over 1,000 data sets in total, we roughly estimate that developmental phonagnosia might occur with 2‰ in the German-speaking population. Estimating prevalence of congenital cognitive deficits is difficult and contentious [31, 32]. In other congenital cognitive deficits, estimates are based on cutoff values in diagnostic tests (e.g., amusia [33]) or on interviews in samples of specific populations (e.g., prosopagnosia [4]). It has been suggested that a combination of such methods yields better estimates [32]. Here, we combined multiple methods, including a web-based test, a questionnaire, an interview, and a laboratory test battery. Nevertheless, we cannot exclude the possible influence of a sampling bias, which could have an effect in either direction (see Supplemental Discussion). We speculate that the true prevalence of phonagnosia is probably higher than 2‰ because the return rate of our four-stage screening approach was relatively low, especially for the questionnaire (see Supplemental Discussion).

General Discussion

In the present study, we demonstrated the existence of developmental phonagnosia as a modality-specific person-recognition deficit in otherwise healthy individuals. This provides support for a central assumption of current models of person recognition, which is that voice recognition can be dissociated from speech, face, and emotion recognition [7, 34].

Although AS and SP performed worse than controls on most of our voice-recognition tests, there was one test for each of them that only showed a trend toward significance. For AS, the voice-face test seemed to be easier. Previous work has shown that simultaneous presentation of the speakers' faces during voice learning generally improves subsequent voice recognition [35–37] and that this improvement is variable interindividually [35]. We speculate that AS received a greater benefit from face information than her controls because she has developed more efficient strategies for using face information to aid her voice-recognition performance. For SP, the results of the voice-name test were vitiated because it transpired that he was familiar with one of the speakers from the test (his former disc jockey teacher), although he did not realize this at the time of testing. Therefore, SP could have been better at this task because of his prior experience with one of the voices (see Supplemental Experimental Procedures).

Many different species readily identify the vocal calls of conspecifics, and the specialization of cortical regions for

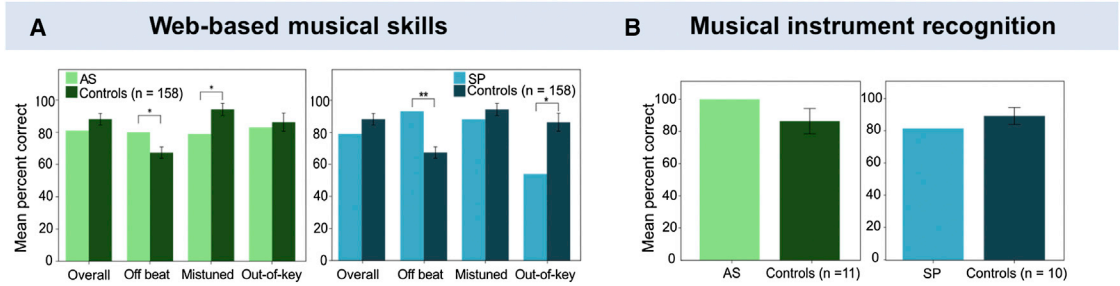


Figure 4. Performance on Tests of Musical Skills

(A) Web-based test of musical skills (http://www.brams.umontreal.ca/onlinetest/). Mean percent corrects of the overall task performance and for each subtest separately are displayed. Asterisks indicate statistical significant differences between AS and SP and control data published in Peretz et al. [26] (*$p < 0.05$, **$p < 0.001$).

(B) Musical instrument recognition test. Mean percent corrects of AS, SP, and their control groups are displayed. There were no group differences on this test.

All error bars show 1 SD. See also Table S7.

voice identification seems to be similar across species [1, 3] and emerges at an early stage of development [38]. Our study raises several new questions about this evolutionary conserved process. For instance, future research may reveal whether phonagnosia is similarly hereditary as prosopagnosia is [39] and how it relates to other auditory cognitive deficits, such as pitch-perception deficits and amusia. Knowledge about these behavioral variabilities in healthy individuals will enhance the understanding of communication idiosyncrasies across different individuals. In addition, because of the selectiveness of the impairment and intact brain structure, developmental phonagnosia provides a unique window into investigating the neuronal mechanisms of auditory person perception.

Supplemental Information

Supplemental Information includes Supplemental Discussion, Supplemental Experimental Procedures, one figure, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.08.048.

Authors Contributions

C.R., S.R.M., F.H., J.K., S.S., and K.v.K. designed the experiment. C.R. performed the experiment and analyzed the data. C.R., S.R.M., and K.v.K. wrote the manuscript.

Acknowledgments

We are grateful to AS and SP for their effort and for the time they took to participate in the extensive test battery and interviews. We thank Stefan Kiebel and Sonja Schall for helpful discussions. We thank Björn Herrmann and Molly Henry for providing the notched-noise test, and we thank Björn Herrmann additionally for help with analyzing the notched-noise test results. Further thanks go to Marc Bangert for implementing the web-based test, to Jason Warren for providing musical instrument stimuli, and to Beate Wendt for providing stimuli used in the vocal-emotion test. We thank three anonymous reviewers for helpful suggestions and comments. This work was funded by a Max Planck research group grant to K.v.K.

Received: April 12, 2014
Revised: July 28, 2014
Accepted: August 20, 2014
Published: September 25, 2014

References

1. Perrodin, C., Kayser, C., Logothetis, N.K., and Petkov, C.I. (2011). Voice cells in the primate temporal lobe. Curr. Biol. *21*, 1408–1415.
2. Sidtis, D., and Kreiman, J. (2012). In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication. Integr. Psychol. Behav. Sci. *46*, 146–159.
3. Andics, A., Gácsi, M., Faragó, T., Kis, A., and Miklósi, Á. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr. Biol. *24*, 574–578.
4. Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., and Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). Am. J. Med. Genet. A. *140*, 1617–1622.
5. Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J.R., Schweinberger, S.R., Warren, J.D., and Duchaine, B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia *47*, 123–131.
6. Van Lancker, D.R., and Canter, G.J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. Brain Cogn. *1*, 185–195.
7. Belin, P., Fecteau, S., and Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. *8*, 129–135.
8. Neuner, F., and Schweinberger, S.R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. Brain Cogn. *44*, 342–366.
9. Hailstone, J.C., Crutch, S.J., Vestergaard, M.D., Patterson, R.D., and Warren, J.D. (2010). Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia *48*, 1104–1114.
10. Perrachione, T.K., Del Tufo, S.N., and Gabrieli, J.D. (2011). Human voice recognition depends on language ability. Science *333*, 595.
11. Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia *9*, 97–113.
12. Crawford, J.R., and Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. Clin. Neuropsychol. *12*, 482–486.
13. Gilaie-Dotan, S., Saygin, A.P., Lorenzi, L.J., Egan, R., Rees, G., and Behrmann, M. (2013). The role of human ventral visual cortex in motion perception. Brain *136*, 2784–2798.
14. Macmillan, N.A., and Creelman, C.D. (2004). Detection Theory: A User's Guide (New York: Psychology Press).
15. Duchaine, B., and Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia *44*, 576–585.
16. Buchtel, H.A., and Stewart, J.D. (1989). Auditory agnosia: apperceptive or associative disorder? Brain Lang. *37*, 12–25.
17. Smith, D.R.R., Patterson, R.D., Turner, R., Kawahara, H., and Irino, T. (2005). The processing and perception of size information in speech sounds. J. Acoust. Soc. Am. *117*, 305–318.
18. Gaudrain, E., Li, S., Ban, V.S., and Patterson, R.D. (2009). The role of glottal pulse rate and vocal tract length in perception of speaker identity. Interspeech 2009: 10th Annual Conference of the International Speech Communication Association 2009 1–5, 152–155.

19. Lavner, Y., Gath, I., and Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. Speech Commun. *30*, 9–26.
20. Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008, 3933–3936.
21. Fitch, W.T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J. Acoust. Soc. Am. *106*, 1511–1522.
22. McDermott, J.H., and Oxenham, A.J. (2008). Music perception, pitch, and the auditory system. Curr. Opin. Neurobiol. *18*, 452–463.
23. Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., and Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. Neuron *33*, 185–191.
24. Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., and Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain *136*, 1639–1661.
25. Marin, M.M., Gingras, B., and Stewart, L. (2012). Perception of musical timbre in congenital amusia: categorization, discrimination and short-term memory. Neuropsychologia *50*, 367–378.
26. Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L.L., Gagnon, B., Trimmer, C.G., Paquette, S., and Bouchard, B. (2008). On-line identification of congenital amusia. Music Percept. *25*, 331–343.
27. Peretz, I., Champod, A.S., and Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. Ann. N Y Acad. Sci. *999*, 58–75.
28. Mathias, S.R., Micheyl, C., and Bailey, P.J. (2010). Stimulus uncertainty and insensitivity to pitch-change direction. J. Acoust. Soc. Am. *127*, 3026–3037.
29. Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., and Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: a problem of recognition. J. Psychosom. Res. *65*, 445–451.
30. Hyde, K.L., and Peretz, I. (2004). Brains that are out of tune but in time. Psychol. Sci. *15*, 356–360.
31. Bowles, D.C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C.E., and Yovel, G. (2009). Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. Cogn. Neuropsychol. *26*, 423–455.
32. Henry, M.J., and McAuley, J.D. (2010). On the prevalence of congenital amusia. Music Percept. *27*, 413–418.
33. Kalmus, H., and Fry, D.B. (1980). On tune deafness (dysmelodia): frequency, development, genetics and musical background. Ann. Hum. Genet. *43*, 369–382.
34. Young, A.W., and Bruce, V. (2011). Understanding person perception. Br. J. Psychol. *102*, 959–974.
35. von Kriegstein, K., Dogan, O., Grüter, M., Giraud, A.L., Kell, C.A., Grüter, T., Kleinschmidt, A., and Kiebel, S.J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. Proc. Natl. Acad. Sci. USA *105*, 6747–6752.
36. Schall, S., Kiebel, S.J., Maess, B., and von Kriegstein, K. (2013). Early auditory sensory processing of voices is facilitated by visual mechanisms. Neuroimage *77*, 237–245.
37. Sheffert, S.M., and Olson, E. (2004). Audiovisual speech facilitates voice learning. Percept. Psychophys. *66*, 352–362.
38. Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G.J., Renvall, V., Deoni, S., et al. (2011). Early specialization for voice and emotion processing in the infant brain. Curr. Biol. *21*, 1220–1224.
39. Kennerknecht, I., Pluempe, N., and Welling, B. (2008). Congenital prosopagnosia—a common hereditary cognitive dysfunction in humans. Front. Biosci. *13*, 3150–3158.

Current Biology, Volume *24*

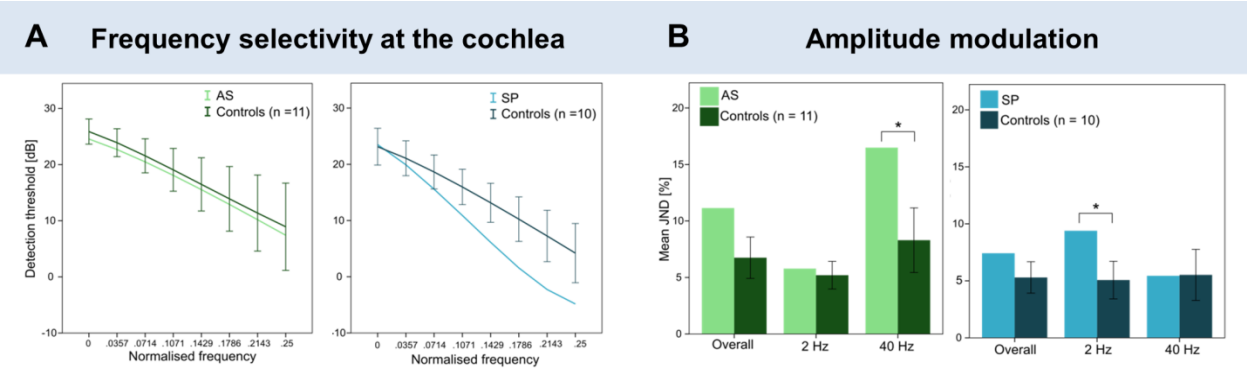
Supplemental Information

Two Cases of Selective Developmental Voice-Recognition Impairments

Claudia Roswandowitz, Samuel R. Mathias, Florian Hintz, Jens Kreitewolf, Stefanie Schelinski, and Katharina von Kriegstein

Supplemental Data

Figure S1. Results of AS and SP and their control groups on the notched-noise and the amplitude modulation tests (see also Table S6), related to Figure 3B/C.



A. Notched-noise test. Mean predicted detection thresholds for the eight normalised frequencies are displayed. A sharper slope depicts higher frequency selectivity. Error bars represent one standard deviation. SP' slope indicates very high frequency selectivity at the cochlea. **B.** Amplitude modulation test. Mean JNDs, just noticeable differences, in amplitude modulation rates are displayed. Error bars represent one standard deviation. Asterisks indicate statistical significant differences between AS and SP and their respective control groups (*$p < .05$).

Table S1. Results for the voice-name learning test, related to Figure 1A/B.

Voice-name learning	Number of participants	Performance		
		Overall Mean [%] (SD)	**Female voice** Mean [%] (SD)	**Male voice** Mean [%] (SD)
Web-based test				
All participants	1057	70.54 (13.64)	60.57 (16.38)	80.63 (15.83)
Age [yrs]				
<18	44	71.89 (13.08)	64.02 (15.82)	79.77 (14.83)
18 - 32	539	74.89 (12.14)	64.90 (15.92)	84.90 (13.80)
33 - 50	313	68.99 (12.52)	58.43 (15.52)	79.55 (14.73)
51 - 67	133	61.05 (12.08)	50.43 (13.72)	71.68 (17.11)
>67	28	49.22 (10.71)	44.05 (10.00)	54.41 (15.37)
Gender				
Female	657	71.44 (12.92)	61.97 (16.05)	80.94 (14.91)
Male	399	69.21 (14.41)	58.28 (16.67)	80.13 (17.26)
Language				
German	1020	70.53 (13.53)	60.37 (16.38)	80.68 (15.80)
Other languages	37	72.50 (13.82)	66.22 (15.32)	79.28 (16.95)
Self-estimation				
Good	574	72.12 (13.25)	61.71 (16.75)	82.23 (14.99)
Don't know	436	69.38 (13.31)	59.88 (15.55)	78.94 (16.04)
Poor	47	63.23 (15.69)	53.19 (17.25)	73.26 (19.96)
Test Repetitions				
0	1029	70.66 (13.48)	60.67 (16.33)	80.66 (15.70)
1	24	69.10 (16.04)	58.06 (18.98)	80.14 (21.09)
2	1	75.00	63.33	86.67
3	2	65.83 (0.83)	48.33 (2.36)	83.33 (14.14)
4	1	45.00	40.00	50.00
≥5	0	-	-	-
Laboratory-based test				
All	42	68.06 (17.00)	59.44 (17.92)	76.67 (19.96)
Age [yrs]				
Mean	43.95			
Range	26 - 71			
Gender				
Female	25			
Male	17			

The table displays mean percent correct and for the control groups additional standard deviations (SDs) on the voice-name learning test. The performance is listed as overall performance and separately for the female and male voice sections of the test. Results from the web-based test are grouped according to the descriptive information participants provided at the beginning of the test: age, gender, native language, self-estimation on voice recognition abilities, and test repetitions. The mean performance of the web-

based test group was slightly higher than of the lab-based test group, but the difference was not significant (permutation, 10000 resamples without replacement; CI = (-4.25, 4.42), $aveDiff_{obs}$ = 3.27, p = .08).

Table S2. Neuropsychological and audiometric assessment of AS and SP

	AS Mean (Percentile)	**SP** Mean (Percentile)
Neuropsychological tests		
TAP Alertness		
Intrinsic [ms]	218 (73)	182 (96)
Phasic [ms]	217 (54)	187 (79)
Vocabulary (WST)		
Hit rate /42	34 (76)	35 (82)
Digit span		
Forwards /12	7 (25)	9 (68)
Backwards /12	5 (17)	8 (80)
Spatial span		
Forwards /14	8 (32)	10 (90)
Backwards /12	8 (32)	9 (55)
Face-name (GNL)		
Overall learning rate /32	30 (58)	29 (29)
Audiometry		
BEA [dB HL]	11.67	8.33
WEA [dB HL]	11.67	16.11
Controls	n = 11	n = 10
BEA [dB HL] (SD)	11.87(4.2)	12.17 (5.83)
WEA [dB HL] (SD)	14.19 (4.25)	14.28 (6.22)

Neuropsychological test results are displayed as mean scores and percentiles. The results of the audiometry are displayed in relation to the control group. AS and SP performed in the normal range on all tests.

TAP: Test of Attentional Performance Version 2.2; [S1], WST: German Vocabulary Test, 'Wortschatztest' – 1st Edition [S2], Digit and Spatial Span [WMS-R; S3], GNL: Face-name learning, 'Gesichter-Namen-Lerntest' [S4]. For all tests a performance at the 2nd percentile would indicate a severe neuropsychological impairment. Hearing levels (HL) are displayed as better ear average (BEA) and worse ear average (WEA). AS and SP's hearing levels at any frequency were equal or below 20 dB and they were not different from the hearing levels of the control group (AS: BEA t(10) = -.05, p = .97, WEA t(10) = -.57, p = .58; SP: BEA t(9) = -.63, p = .55, WEA t(9) = .28, p = .79).

Table S3. Results of AS and SP and their control groups on all voice-processing tests, related to Figure 2A/B, 3A.

Voice tests	**AS** Mean	**Controls** Mean (SD)	**ΔSDs**	**t value**	**p value**	**SP** Mean	**Controls** Mean (SD)	**ΔSDs**	**t value**	**p value**
Unfamiliar voice learning										
Voice-name										
Overall [%]	50.00	74.97 (11.64)	2.15	-2.14	.016*	55.00	74.90 (12.97)	1.53	-1.53	.064
Female [%]	46.67	65.11 (15.70)		-1.17	.121	36.67	63.91 (16.49)		-1.65	.051
Male [%]	53.33	84.82 (12.90)		-2.44	.008*	73.33	84.89 (14.71)		-.78	.217
Voice-colour										
Overall [%]	46.67	74.15 (8.53)	3.33	-3.09	.006*	47.33	76.72 (5.58)	5.27	-5.02	<.001**
Female [%]	50.67	72.64 (9.83)		-2.14	.029	45.33	68.53 (12.15)		-1.82	.051
Male [%]	42.67	74.16 (10.15)		-3.12	.006*	49.33	83.84 (8.68)		-3.79	.002*
Voice-face										
Overall [%]	73.33	86.79 (7.82)	1.72	-1.64	.067	67.34	89.60 (2.37)	9.39	-8.96	<.001**
Female [%]	57.33	84.82 (10.63)		-2.47	.018*	53.00	84.80 (5.60)		-5.59	<.001**
Male [%]	89.33	88.75 (8.15)		.068	.47	82.67	94.40 (3.47)		-3.23	.005*
Famous voice recognition										
Familiarity decision										
c	-1.5	-.22 (.31)	4.13	-3.92	.003*	1.09	-.07 (.39)	4.59	2.85	.02*
d'	.19	2.11 (.5)	3.84	-3.68	.002*	1.49	2.05 (.66)	.85	-.81	.22
Voice naming										
Correct naming [%]	88.89	80.97 (16.45)	.48	.46	.327	36.67	82.1 (4.22)	10.76	-10.26	<.001**
Voice discrimination										
Overall [%]	68.52	96.64 (3.22)	8.73	8.47	<.001**	90.74	94.90 (5.38)	.77	-.75	.47
c	-.92	-.01 (.19)	4.79	-4.78	<.001**	-0.19	-.04 (.18)	.83	-.81	.43
d'	1.98	3.65 (.43)	3.88	-3.73	<.001**	2.81	3.43 (.55)	1.13	-1.08	.30

The table displays mean scores and for the control groups additionally the standard deviations (SDs). ΔSDs indicate the standard deviations that AS and SP performed below their controls. Controls refers to the respective control groups of AS and SP for the web-based test and the laboratory tests (see text and

Supplemental Experimental Procedures). Significant differences between AS/SP and their control groups are indicated by asterisks (*p < .05, **p < .001). Note that SP's poor voice-naming performance cannot be explained by a general naming deficit because he performed well on other non-auditory naming tests (see neuropsychological assessment in Table S2 and 'face-name learning test' in Table S4).

Table S4. Results of AS and SP and their control groups on all control tests, related to Figure 2C.

Control tests	**AS** Mean	**Controls** Mean (SD)	**t value**	**p value**	**SP** Mean	**Controls** Mean (SD)	**t value**	**p value**
		n = 11				n =10		
Speech-in-noise SRT [dB]	-8.77	-9.18 (1.58)	.25	-809	-9.76	-8.80 (1.54)	-.59	.567
Vocal-emotion [%]	82.50	83.56 (6.41)	-.16	.877	81.67	81.34 (9.0)	.04	.973
CFMT [%]	87.50	79.67 (10.02)	.75	.472	80.56	80.69 (12.94)	-.010	.993
Face-name learning [%]	96.67	86.06 (10.73)	.95	.366	96.00	86.47 (15.76)	.577	.578

The table displays mean scores and for the control groups additionally the standard deviations (SD). Speech-reception thresholds (SRTs) measure the signal-to-noise ratio in dB.

Table S5. Results of AS and SP and their control groups on the vocal pitch and vocal timbre discrimination test and additional auditory abilities, related to Figure 3B/C, Figure S1.

Acoustic voice tests	**AS** Mean	**Controls** Mean (SD)	**ΔSDs**	**t value**	**p value**	**SP** Mean	**Controls** Mean (SD)	**ΔSDs**	**t value**	**p value**
		n = 11					n = 10			
Vocal pitch JND [cent]	170.08	49.13 (39.20)	3.09	2.95	.014*	137.2	32.17 (26.10)	4.02	3.84	.004*
Vocal timbre JND [SER]	5.43	4.75 (2.61)	.26	.25	.81	5.34	4.13 (3.34)	.36	.35	.74
Additional auditory abilities										
Notched noise										
Predicted threshold [dB]	16.48	17.63 (4.18)	.28	-.26	.8	8.81	14.21 (3.15)	1.71	-1.64	.14
Pass-band p	20.93	22.4 (6.75)	.22	-.21	.84	37.61	22.74 (5.67)	2.62	2.5	.034*
Amplitude modulation JND [%]										
Overall	11.14	6.75 (2.39)	1.84	1.76	.11	7.42	5.29 (1.38)	1.54	1.47	.18
2 Hz	5.78	5.2 (1.22)		.46	.66	9.39	5.07 (1.65)		2.5	.03*
40 Hz	16.49	8.3 (2.73)		2.87	.02*	5.44	5.52 (2.24)		-.03	.97

The table displays mean scores and for the control groups additionally the standard deviations (SDs). ΔSDs indicate the standard deviations that AS and SP performed below their controls. Significant differences between AS/SP and their respective control groups are indicated by asterisks (*$p < .05$, **$p < .001$). Cent describes the logarithmic unit for F0 intervals. SER: spectral envelope ratio. Spatial envelope is a unit for the acoustic effect of the speaker's vocal-tract length. Pass-band p: the slope of the auditory filter. A higher value depicts higher frequency selectivity. SP slope of the auditory filter was significantly sharper than the slope of his control group indicating a very high frequency selectivity at the cochlea. For subtest-specific discussion on the AM modulation test see Supplemental Discussion.

We tested for correlations between the pitch JNDs and the overall voice-learning performance of all controls. There was no significant correlation ($r(21) = -.01$, $p = .96$) indicating that the ability to discriminate vocal pitch does not necessarily predict voice-recognition abilities.

Table S6. Structured interview to assess musical abilities of AS and SP

	AS	SP
1. *How would you rank your ability to detect when someone else sings out of tune?** *(excellent – good – fair – poor – very poor)*	good	good
2. *How would you rank your ability to remember very familiar melodies without the help of the lyrics?** *(excellent – good – fair – poor – very poor)*	excellent	good
3. *How often have you been told to sing out of tune?* *(never – seldom – sometimes – often – very often)*	sometimes	often
4. How often do you listen to music? (very often – often – sometimes – seldom – never)	seldom	often
5. Do you like listening to music? (strongly agree – agree – neither agree or disagree – disagree – strongly disagree)	agree	agree
6. How would you rank your ability to sing back a note played on the piano? (excellent – good – fair – poor – very poor)	fair	very poor
7. How often do you sing in private? (very often – often – sometimes – seldom – never)	seldom	seldom
8. How often do you sing in public? (very often – often – sometimes – seldom – never)	seldom	never
9. Do you like dancing? (strongly agree – agree – neither agree or disagree – disagree – strongly disagree)	agree	agree
10. How often do you dance? (very often – often – sometimes – seldom – never)	sometimes	often

We interviewed AS and SP with a questionnaire that has been previously used in congenital amusia research [S5]. The first three questions are key questions for diagnosing amusia. The first two have a high specificity (indicated with an asterisk); 85% of amusics report that they are unable to detect when someone sings out-of-tune, but none of the controls and 74% of amusics report that they are poor at recognising familiar melodies without lyrics, but just 8% of the controls. The third question has a high sensitivity, but low specificity, because 93% of amusics report that they have been told to sing out of tune, but also 43% of the controls [S5].

Table S7. Results of AS and SP and their control groups on all musical-processing tests, related to Figure 4A/B.

Musical-processing tests	**AS** Mean	**SP** Mean	**Controls** Mean (SD)	**AS t value**	**p value**	**SP t value**	**p value**	
Web-based musical skills			n = 158					
Overall [%]	81	79	88.1 (7.2)	-.98	.327	-1.26	.21	
Off-beat /30	24	28	20.20 (2.1)	1.80	.073	3.70	> .001**	
Mistuned /24	19	21	22.6 (1.8)	-1.99	.048*	-.89	.377	
Out-of-key /24	20	13	20.7 (2.7)	-.26	.796	-2.84	.005*	
Musical instrument recognition	**AS** Mean	**Controls** Mean (SD)	**t value**	**p value**	**SP** Mean	**Controls** Mean (SD)	**t value**	**p value**
		n = 11				n = 10		
Overall [%]	100	86.34 (7.87)	1.66	.128	81.82	89.54 (5.27)	-1.40	.196

The table displays mean scores and for the control groups additionally the standard deviations (SD). Controls refer to AS and SP's respective control groups for the web-based musical test (published in Peretz et al. [S5]) and the laboratory test (see Supplemental Experimental Procedures). The amusia cut-off value for the overall performance on web-based musical skill test is 73.7%. Significant differences between AS/SP and their respective control groups are indicated by asterisks (*$p < .05$, **$p < .001$). The overall musical performance of AS and SP was in the normal range. AS and SP performed both better than the controls on the off-beat test, but worse than controls on one of the two pitch-based tests.

Supplemental Experimental Procedures

Participants

Cases AS and SP

AS studied psychology and first learned about phonagnosia when she prepared a student report about prosopagnosia. She realised that she might have a similar deficit in the auditory modality, and she came across our web-based test via an internet search on auditory person-identity recognition deficits.

SP is local and took part in a PhD course together with the first author. While talking about their respective PhD topics, SP became interested in participating in the web-based test. Both AS and SP reported no neurological or psychiatric disorders and were free of medication. As a child, SP suffered from developmental stuttering but with speech therapy he recovered from it. AS played an instrument (recorder) for four years. SP reported that he plays the guitar since 15 years at an amateur level and had professional DJ lessons for 1.5 years. AS and SP interviewed close family members and none of them reported voice-recognition difficulties.

AS and SP gave their informed written consent prior to testing according to procedures approved by the Research Ethic Committee of the University Leipzig and were paid for their participation.

Further cases invited to the lab

The other two individuals that we had tested in the laboratory based experiment had to be excluded. One participant had weak performances in the neuropsychological assessment and the other had to be excluded because of a history of head injury with coma that only transpired when screening him for the exclusion criteria for MRI scanning.

Control participants

Web-based controls matched to AS and SP

Web-based control participants were matched to AS and SP in age and gender. The control group of AS comprised 343 age matched (age range = 18 - 32 years) native German female participants.172 native German male participants within the age range of 18 - 32 years were included into SP's control group.

Laboratory controls for the web-based voice-name test

We collected data of 42 participants (mean age = 43.07 years, SD = 14.18, age range = 26-71 years, 28 female). All participants gave their informed written consent prior to testing according to procedures approved by the Research Ethic Committee of the University Leipzig. All were paid for their participation in the study (7€/h).

Laboratory controls matched to AS and SP

Laboratory control participants were matched to AS and SP in the following criteria: age, gender, education (university entrance diploma), and handedness (Edinburgh questionnaire; [S6]). None of the controls reported a history of neurological disease and/or deficits in person-identity recognition. AS control group consisted of 11 female controls (mean age = 32 years, SD = 0.6, range = 31 - 33 years) for the famous voice-recognition test, the vocal-pitch and vocal-timbre test, the vocal-emotion, musical instrument, speech-in-noise test, and the Cambridge Face Memory Test, 11 female controls (mean age = 30.72 years, SD = 1.21, range = 29 - 33 years) for the voice-colour, voice-face, and face-name test and 11 female controls (mean age = 32.1 years, SD = 0.67, range = 31-33 years) for the notched noise and amplitude modulation test. This was the case, because five controls could be re-invited for the second and five for the third and fourth testing session. The control group of SP contained ten male participants (mean age = 32.18 years, SD = 0.4, range = 32 - 33 years) for the laboratory tests, except the notched noise and the amplitude modulation test. In these, one of the previous participants could not be re-invited, so that we had to include a new participant (mean age = 31.8 years, SD = 0.75, range = 31 - 33 years). For the voice-discrimination test, the control group of AS and SP contained each 16 participants (AS's controls: all female, mean age = 31.63 years, SD = 1.22, range = 30 - 34 years; SP's controls: all male, mean age = 31.06 years, SD = 1.2, range = 29 - 33 years). We included 16 participants in this test, because it was part of a different study that also included functional MRI, which necessitates larger subject groups. All participants gave their informed written consent prior to testing according to procedures approved by the Research Ethic Committee of the University Leipzig and were paid for their participation in the study.

Experimental procedure

Web-based voice-name test

We undertook an extensive media campaign to bring the web-based test to public attention, involving radio interviews, newspaper articles, and public-science lectures.

Stimuli and presentation software

The stimuli were high-quality speech recordings made from six native speakers of standard German (three female, age range = 22 - 29 years). All speakers were students of speech sciences and had received professional voice and speech training as part of their study program. We instructed the speakers to read German sentences at a normal speech rate with an emotionally neutral intonation. Each of the six speakers read 26 five-word declarative sentences (e.g., German: "Der Junge trägt einen Koffer.", English: The boy carries a suitcase.), 3 two-word declarative sentences (e.g.,"Er sagt.", He says.), and 3 five-word interrogative sentences ("Trägt der Junge einen Koffer?", Does the boy carry a suitcase?).

The stimuli were recorded in a sound-attenuating chamber using a condenser microphone (Neumann TLM 50, Berlin; Mic-Peramp: Mic-Amp F35, Lake People, Germany; Soundcard: Power Mac G5 Dual 1.8 GHz, Apple Inc., CA, USA; 44.1 kHz sampling rate, and 16 bit resolution). They were recorded using Sound Studio 3 (Felt Tip Inc., NY, USA), post-processed using Audacity (version 1.3.5. beta (http://audacity.sourceforge.net) and Matlab (version 8.1, The MathWorks, Inc., MA, USA), and were normalised for peak amplitude using PRAAT [S7]. The test was written in Adobe Flash (version 11; Adobe Systems, Inc., San Jose, CA, US) and responses were recorded via a computer mouse. The test is assessable at http://www.phonagnosie.de.

Procedure

The web-based test first asked participants to provide demographic information, including their age range (< 18, 18 - 32, 33 - 50, 51 - 67, > 65 years), gender (female, male), mother tongue, and the number of times they had previously taken the test (0, 1, 2, 3, 4, >5). It also recorded their email address, and how they would estimate their own voice-recognition ability (good, don't know, poor).

The test contained a 'female-voices' block and a 'male-voices' block, which were identical in structure. In each block, participants learned to associate three unfamiliar voices with three first names. The order of blocks was randomised between participants. The blocks comprised four learning and four testing phases, completed in alternating order. During the learning phases, vocal stimuli were heard whilst the associated name of the speaker (Anna, Laura, Ruth for the females and Felix, Paul, Moritz for the males voices) was printed to the screen simultaneously. The four learning phases had different numbers of five-word declarative sentences: four sentences per speaker were heard in the first phase; three in the second; and two in the third and fourth phases. The running order of speakers was randomised within all learning phases. During the test phases, the participants listened to vocal stimuli, and after each stimulus, they selected the speaker's name from the three alternatives. The first and second testing phases contained feedback: after each trial, the correct name of the speaker was presented and the stimulus was heard again. Feedback was turned off for the third and fourth testing phases. The four testing phases also differed in the number and syntactical structure of the presented vocal stimuli. During the first and second phases, two five-word declarative sentences were heard per speaker. During the third phase, three five-word interrogative sentences were heard per speaker. In the fourth phase, three two-word declarative sentences were heard per speaker. Participants were guided through the test by written instructions. At the end of the test, individual test results were presented on the screen and the data were transferred to us via email. The whole test could be completed within approximately 20 minutes.

Questionnaire

We developed a paper-based voice-recognition questionnaire that asked the participants to rate their own abilities on everyday voice-related tasks (e.g., identify a speaker's gender or accent; follow a radio interview; etc.), as well as their experiences with general person recognition, general auditory perception,

and medical history. An English version of the questionnaire is available online (http://kriegstein.cbs.mpg.de/questionnaire). We selected participants for the telephone interview if they reported: (i) voice-recognition difficulties for personal familiar and/or famous voices, (ii) no difficulties in understanding speech perception in a quiet environment, (iii) no difficulties with extracting other vocal information from the voice (i.e. gender, age, emotion, dialect), and (iv) normal music appreciation.

Telephone interview

During the semi-structured telephone interviews, we checked for any history of neurological or psychiatric illnesses. In addition, we asked the participants to describe in more detail their voice-recognition impairments, for instance, by providing specific situations in which they realised they had difficulties with voice recognition. We also asked the participants for details about the environment in which they had performed the web-based test, e.g. headphones or loudspeakers, noisy or quiet background. We selected participants for the laboratory tests, if they met the following criteria: No reported history of neurological or psychiatric illness; ability to report convincing details about the voice-recognition impairment.

Laboratory experiment

Audiometry

We assessed hearing levels via pure-tone audiometry using a screening audiometer (MADSEN Micromate 304, GN Otometrics, Copenhagen, Denmark). Hearing level was assessed in each ear at frequencies separated by octave steps ranging from 250 to 8000 Hz [S8].

Neuropsychological assessment

To assess the cognitive abilities of AS and SP, we conducted a battery of neuropsychological tests that assessed attention and alertness (Test of Attentional Performance Version 2.2; [S1]), verbal-intelligence and language-comprehension skills (German Vocabulary Test, 'Wortschatztest' – 1st Edition [WST; S2]), auditory and spatial working-memory capacities (Digit and Spatial Span [WMS-R; S3], and associative-learning and naming abilities ('Face-name learning' [GNL; S4]).

Structural MRI

MR images of AS and SP were acquired using a 3-Tesla Magnetom Verio scanner (Siemens, Erlangen, Germany) using a 32-channel head coil. The protocol included a T1-weighted MPRAGE sequence (TR = 2.3 ms; TE = 2.98 ms; TI = 900ms; flip angle = 9°; FOV = 256 × 240 mm; 176 sagittal slices; voxel size = 1 × 1 × 1 mm) and a 3D T2-weighted FLAIR sequence (TR = 10000 ms; TE = 90 ms; TI = 2500 ms; flip angle = 180°; FOV = 220×165 mm; 26 sagittal slices; voxel size = 0.9 × 0.9 × 4.0 mm). An experienced neurologist inspected the sMRI and FLAIR images of AS and SP for abnormalities. There were several

white matter hyperintensities in bilateral frontal lobes in AS that were rated as unspecific by an experienced neuroradiologist. There were no structural abnormalities in SP's brain scan.

Laboratory test battery

General procedure

All tests, except for the voice-discrimination, notched noise, and the amplitude modulation test, were carried out on a desktop computer. Participants were comfortably seated facing a 21-inch monitor that displayed the visual stimuli. The voice-discrimination, the notched noise, and the amplitude modulation test were carried out on a laptop computer facing a 14.1-inch monitor. Auditory stimuli were presented via Headphones (Sennheiser HD 280 pro, Wennebostel, Germany). The sound level was adjusted to a comfortable sound pressure level (SPL) for each participant (average SPL: approximately 70 dB). The participants' responses were either recorded via a keyboard or a custom-made response box, depending on the test. To ensure that all participants understood the tasks, the experimenter gave oral instructions in addition to the written instructions prior to each test.

We tested AS in four separate sessions. The first and the second session were carried out in the lab and lasted around 2.5 hours. We conducted the third and fourth sessions in a quiet room at her home due to organisational reasons. These sessions lasted around one hour each. Control participants were tested in the lab. One control who participated in the first and second session had to be excluded from the analysis because she reported voice-processing difficulties in her everyday life and performed 1.23 SDs below normative controls reported in Duchaine et al. [S9] in the standardised face-recognition test (CFMT), indicating additional face-recognition deficits.

Due to personal time constraints that did not allow SP to come to the lab for the longer testing sessions (first and second), we distributed the testing over four shorter sessions. SP's control group was tested over two sessions, each lasting 2.5 hours. For SP and his controls, we conducted the third and fourth session in the lab; each lasting around one hour.

Voice-processing tests

Unfamiliar voice-learning tests: 'Voice-colour' and 'Voice-face'

Since the web-based test required participants to associate voices with written names, deficits in name processing (or in matching voices to names) could have caused poor performance. To determine whether AS and SP suffered from name-processing deficits, we included two additional unfamiliar voice-learning tests in the laboratory battery, which required participants to associate unfamiliar voices with two different kinds of visual stimuli, colours and faces.

Stimuli and presentation software

The auditory stimuli were recorded from 14 native speakers of standard German (eight female, age range = 21 - 32 years). We instructed all speakers to read the sentences with a normal speech rate in an emotionally neutral intonation. We used the same sentence types as in the web-based test. Each speaker read 41 five-word declarative sentences, 5 two-word declarative sentences, and 5 five-word interrogative sentences. This resulted in a total set of 714 sentences. High quality auditory recordings were taken in a soundproof recording chamber with a condenser microphone (Rode NT 55 MP; USB Sound Interface: Fast Track MK2, M-Audio, US; 44.1 kHz sampling rate, 16 bit resolution) and Audacity software (version 1.3.5. beta (http://audacity.sourceforge.net). The stimuli post-processing was the same as for the stimuli of the web-based test. The visual stimuli comprised photographic images of the speakers' faces. Images were recorded with a digital video camera (Legria HF S10 HD-Camcorder, Canon Inc., Japan); all taken under the same lighting conditions in front of a black background. The speakers' faces were visible from the chin to the hairline with a neutral expression. No face contained salient visual features such as beards, piercings or glasses. The test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Procedure

The procedure of the voice-colour and the voice-face learning test were exactly the same, although different visual stimuli and different speakers were used in both tests. When testing SP (and SP's control participants) on the original voice-face test, two of the female speakers had to be replaced, because SP recognised their faces; one from university and one from his home town. This was pure coincidence. We attribute his acquaintance with these two females and one of the male speaker presented in the voice-name test (his DJ teacher) because he is local and in the same age-range as the speakers.

The tests contained a 'female-voices' block and a 'male-voices' block, which were identical in structure. The blocks were structured into four learning and four testing phases, presented in alternating order, plus an additional final testing phase. During learning, participants attended to the voice-colour/ voice-face pairs. The learning phases differed in the number of five-word declarative sentences that were presented: ten sentences per speaker during the first and second learning phase and three sentences per speaker in the third and fourth learning phases. The running order of the speakers was randomised. During the testing phases, the participants listened to the auditory stimuli and performed a three-alternative forced choice task in which they selected either the colour (voice-colour test) or the face (voice-face test) associated with the voice. In the first and the second testing phases, participants received feedback about their decision, and the correct voice-colour/ voice-face pair was presented again. Each of the speakers contributed five sentences to all of the testing phases, however the syntactical structure altered during testing to avoid prosody-driven identity processing: five-word interrogatives in the third phase and two-word declaratives in the fourth phase. The tests each took approximately 40 minutes to complete.

Famous voice-recognition test

Stimuli and presentation software

The auditory stimulus set contained voice samples of famous (n = 40) and non-famous (n = 20) German speakers. We extracted the voice samples from open-access audio files available on public radio and television websites. Each sample was cut to five seconds duration. The famous voice samples comprised voices from 21 media personalities, eight politicians, seven actors, and four musicians (see below). In a pilot study, a group of 10 individuals without voice-recognition deficits rated the familiarity of a larger stimulus set (n = 56) on a scale from 0 (completely unfamiliar) to 5 (highly familiar). Samples rated with an average of three or higher (n = 40) were included in the final stimulus set. The files were edited using Audacity (version 1.3.5. beta (http://audacity.sourceforge.net)) and peak amplitude was scaled using PRAAT [S7]. The semantic content of the auditory samples provided no information about the famous persons' identity or profession. The test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Names of the famous Germans presented in the famous voice-recognition test (n = 40)

Marcel Reich-Ranicki	Michael Mittermeier	Sabine Christiansen
Hellmuth Karasek	Gerhard Schröder	Anne Will
Franz Beckenbauer	Helmuth Kohl	Alice Schwarzer
Boris Becker	Erich Honecker	Verona Pooth
Dieter Thomas Heck	Norbert Blühm	Heidi Klum
Harald Schmidt	Hans Dietrich Genscher	Barbara Schöneberger
Alfred Biolek	Herbert Grönemeyer	Nena
Thomas Gottschalk	Peter Maffay	Bärbel Schäfer
Jürgen von der Lippe	Udo Lindenberg	Angela Merkel
Oliver Pocher	Martin Semmelrogge	Ulla Schmidt
Harpe Kerkelingen	Jürgen Vogel	Heike Makatsch
Stefan Raab	Ben Becker	Uschi Glas
Otto Walkes	Till Schweiger	Katja Riemann
Helge Schneider		

Procedure

At the beginning of the test, participants were asked to estimate their weekly exposure to television and radio (in hours). The main test comprised two parts. In the first part, participants were presented with the voice samples and were then asked to categorise the voices as familiar or unfamiliar. If on a given trial they categorised the voice as familiar, they subsequently provided that person's name or another characteristic such as their occupation. If the voice sample was categorised as unfamiliar, the next voice

sample was presented. One practice trial was introduced to familiarise participants with the task. In the second part of the test, we assessed each participant's familiarity with the identities of the famous voices they heard earlier. Participants were presented with each famous person's written name (in a different order to the voice samples), and rated how familiar they would find their voice on a scale from 1 to 7 (very unfamiliar to very familiar). The complete test took about 30 minutes.

Analysis

Based on their personal familiarity ratings obtained from the second part of the test, we individually re-classified all of the famous voices (n = 40) into famous and non-famous voice categories. Celebrities' voices rated with at least 5 were classified as famous and celebrities' voices rated with 4 or below were classified as non-famous. The non-famous voices (n = 20) were all classified as non-famous. This procedure allowed us to determine whether the participants correctly or incorrectly categorised any subjective famous voice sample as familiar or unfamiliar in the first part of the test. AS rated 9 famous voices as familiar in the second part (controls: n = 16), and SP rated 30 voices as familiar (controls: n = 23). This procedure in turn allowed us to analyse the data by applying detection theory [S10]. We computed indices of sensitivity (d-prime, *d'*) and response bias (*c*) assuming the yes-no decision model. Each famous voice sample correctly classified as familiar was considered a 'hit', and each famous voice sample incorrectly classified as familiar was considered a 'false alarm'.

In addition to applying detection theory to participants' familiarity decisions, we also computed proportion of those familiar voices who were correctly identified (e.g., by their name or occupation).

Voice-discrimination test

Stimuli and presentation software

The stimulus set consisted of 16 German subject-verb sentences (German: 'Er kauft.', English: 'He buys.'). They were spoken by three male native German speakers (22, 25, and 26 years old). The sentences were semantically neutral, phonologically, and syntactically homogenous and spoken in a neutral manner. None of the speakers had been used in the other voice tests. We recorded the speakers under the same conditions as described earlier (cf. 'voice-name stimuli'). The test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Procedure

The voice-discrimination test contained two parts. In the first part, the participants were shortly familiarised with the voices of each speaker by listening to the sentences spoken by each of the three speakers. The sentences were organised in blocks and each block contained 10 sentences per speaker. In total, there were two blocks per speaker; each presenting the same sentences. The order of the speaker blocks was randomised. In the second part, participants were tested on their voice-discrimination

abilities. To do this, pairs of sentences spoken by the same or two different speakers were presented. Participants indicated whether the sentences were spoken by the same speaker or by different speakers by pressing two buttons labelled as ‘same’ or ‘different’ on the keyboard. Whether the speakers spoke the same or different sentences was randomised. Visual feedback about correct and incorrect responses was provided immediately after each trial. There were 54 trials in total. The test including the familiarisation period lasted approximately 15 minutes.

Analysis

We computed the mean percent correct responses on trials were stimuli pairs were correctly classified as same speakers. In addition, we computed indices of sensitivity (d-prime, *d’*) and response bias (*c*) assuming a yes-no decision model applying detection theory [S10]. Each stimulus pair correctly classified as same speakers were considered as ‘Hit’ and each stimulus pair incorrectly classified as same speakers were considered as ‘False alarm’.

Auditory control tests

Speech-in-noise test

Stimuli and presentation software

The stimulus set consisted of 134 German subject-verb sentences (”Er schreibt.”, He writes.) spoken by a male native German speaker whose voice had not been used in any of the previous tests. We recorded the speaker under the same conditions as described earlier (cf. ‘voice-name stimuli’). The stimuli were mixed with speech-shaped noise to the speech signal using Matlab (version 7.7, The MathWorks, Inc., MA, USA) and Cogent 2000 (www.vislab.ucl.ac.uk/cogent_2000.php). By doing so, the noise was matched to the average long-term spectral properties of the speech signals. Speech-shaped noise has a comparable masking on the speech as adding a number of other speakers talking at the same time (‘multi-speaker babble’), simulating real life noise situations (e.g., a noisy party [S11]). The test was programmed in Matlab (version 7.7, The MathWorks, Inc., MA, USA) and Cogent 2000 (www.vislab.ucl.ac.uk/cogent_2000.php) and responses were recorded via a response button box.

Procedure

The participants listened to one subject-verb sentence per trial and were asked to select the verb out of four alternatives. Among the alternatives was the target verb, and the others had different verbs which were similar to the target verb in their first phonemes. Participants’ individual speech-reception thresholds (SRTs) were measured in five runs of an adaptive-tracking procedure, which used the weighted one-up, one-down method that estimated the signal-to-noise ratio (SNR) in dB at which the participant responded with 75% accuracy [S12]. The initial SNR was 10 dB; this value decreased in steps of 1.3 dB following

each correct response and increased in steps of 4 dB following each incorrect response. After four reversals (a switch from correct to incorrect response or vice versa within two consecutive trials), the up and down step sizes were changed to 2 and 0.7 dB, respectively, and the block of trials continued for a further 10 reversals. SRT was estimated from a single run by taking mean of all SNR values visited during the final 10 reversals, and the participant's overall SRT was defined as the mean SRT from five runs. After each run, the participants could decide on their own when to continue on to the next run by pressing a button. Feedback about correct and incorrect responses was provided after each trial. Before the test started, the participants carried out three practice trials. The test lasted approximately 20 minutes.

Vocal-emotion test

Stimuli and presentation software

The stimulus set consisted of verbal stimuli spoken by one actor and one actress (for detailed stimuli description see SWendt [13]). Auditory stimuli expressed five basic emotions [S14] — sadness, anger, fear, joy, and disgust — and as well as an emotionally neutral state. The sentences were all emotionally neutral in semantic content. Each speaker spoke 10 two-syllable nouns per emotion, so that the complete test material consisted of 120 nouns. All stimuli were peak-amplitude scaled using PRAAT [S7]. The test was implemented in Presentation (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded with a response button box.

Procedure

On each trial, participants heard one noun and responded by selecting the emotional category from the six alternatives (six-alternative forced choice task). The order of stimuli was random. No feedback was given during the test. The test started with an auditory example of each emotional category spoken by both speakers and two practice trials. The test lasted approximately 20 minutes.

Visual control tests

Face-name learning test

Stimuli and presentation software

The stimulus set consisted of images of six male faces (three British and three Spanish actors who are not famous in Germany). The pictures were downloaded from freely accessible websites. The images were degraded by adding different levels of Gaussian noise to each pixel (e.g., 15%, 30%, 60%) using Adobe Photoshop CS4 (version 11.0.2; Adobe Systems, Inc., San Jose, CA, US). The complete stimulus set contains 228 images. Original photographs were taken from different views with varying facial expressions and under different lighting conditions. The test was implemented in the Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via keyboard.

Procedure

The test was structured into four learning and four testing phases, presented in alternating order, plus an additional final testing phase. During the learning phases, participants attended to face-name pairs (Peter, Jan, Timo, Alex, Otto, Leon). The learning phases differed in the number of presented face-name pairs: 10 face-name pairs per person during the first learning phase, three face-name pairs per person during the second, and two face-name pairs per person during the third and fourth learning phase. During the testing phases, the participants were presented with a novel face image and they selected the corresponding name from the six alternatives. Each testing phase contained five items per person. During the second, third, and fourth testing phases, pictures were added with increasing noise levels (15%, 30%, and 60% respectively) to minimise ceiling effects. In the first and second testing phase, participants received feedback on their decision and the correct face-name pair was presented again. The test took approximately 30 minutes.

Acoustic voice-processing tests

Vocal-pitch and Vocal-timbre test

Stimuli and presentation software

The stimulus set consisted of five English vowels (/a/, /e/, /i/, /o/, /u/) resynthesised using the STRAIGHT software package [S15] implemented in Matlab (version 7.7, The MathWorks, Inc., MA, USA). The original vowels were spoken by a male speaker (same material used in [S16]). Prior to resynthesis, the vowels were modified to be monotonic and 600 ms in duration. For the vocal-pitch task, all tokens of a given vowel were identical except for their fundamental frequency. F0 is the physical correlate of a speaker's glottal-pulse rate (GPR), which determines their voice pitch. For the vocal-timbre task, all tokens of a given vowel were identical except for their spectral envelopes, which were scaled proportionally up or down in log-frequency space from the original spectral envelope. Spectral envelope is the physical correlate of a speaker's vocal-tract length (VTL), which is an aspect of vocal timbre that correlates with speaker size [S17]. Both tests were implemented in Python (version 2.7.3, http://python.org/) and responses were recorded via a keyboard.

Procedure

We used an adaptive-tracking procedure [S12] to measure the participants' pitch and timbre just-noticeable differences (JNDs). During the vocal-pitch test, participants listened to pairs of sequentially presented vowels differing only in their F0. One vowel per trial always had an F0 of 112 Hz, and the other was higher in F0 by an amount (ΔF0) defined in musical cents (1 semitone = 100 cents). The order of the stimuli was random on each trial, and participants reported which one was higher in pitch. The initial ΔF0 was 100 cents; this value decreased in steps of 10 cents following each correct response and increased

in steps of 30 cents following each incorrect response. After four reversals (a switch from correct to incorrect response or vice versa within two consecutive trials), the up and down step sizes were changed to 6 and 2 cents, respectively, and the block of trials continued for a further 10 reversals. JND was estimated from a single run by taking mean of all ΔF0 values visited during the final 10 reversals, and the participant's overall JND was defined as the mean JND over five runs. Feedback about response accuracy followed each trial. At the beginning of the test, participants were familiarised with the auditory stimuli by presenting them with two vowels at the extremes of the F0 range. The average test duration was 15 minutes.

The vocal-timbre test was identical to the vocal-pitch test except that the stimuli on each trial differed in their spectral envelopes, and participants reported which vowel was spoken by the smaller speaker. One vowel on each trial had a spectral envelope equal to that of the original speaker, while other differed by ΔSER, defined in percent. Initial ΔSER was 12%; up and down step sizes were 3% and 1% for the first four reversals, and 0.6% and 0.2% for the remaining 10 reversals.

Additional tests of auditory abilities

Notched-noise test

Stimuli and presentation software

To measure individual auditory filter functions, we used a so-called 'notched-noise test' [S18]. This test was kindly provided by the Auditory Cognition group at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, and has been published previously [S19]. With this test, we acquired detection thresholds for a 1400-Hz sine tone in noise. The noise was filtered white noise in which a spectral notch was centred at 1400 Hz between two white noise bands each with a width of 560 Hz. The width of the notch varied parametrically with eight different values (0, 100, 200, 300, 400, 500, 600, 700Hz). For each participant, stimuli were generated using Matlab (version 7.7, The MathWorks, Inc., MA, USA). Based on individual hearing thresholds acquired before the test, the level of the notched noise was fixed at 45 dB above the hearing threshold. The test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via a keyboard.

Procedure

On each trial participants were presented with three notched-noise stimuli. One of them contained the 1400Hz-sine tone. We asked participants to indicate the stimulus in which the tone occurred (i.e. first, middle, or last stimulus). Responses were given via three keys on a keyboard. Feedback about response accuracy was provided after each trial. One block ended after 12 reversals. The test included eight blocks in which stimuli with different notch widths were presented separately. Within one block, the intensity of

the sine tone varied dependent on the participant's performance. We used a one-down one-up staircase procedure to estimate the intensity of the sine tone yielding 50% detection [S20]. Detection thresholds measuring the average tone intensity for each of the eight notch widths were calculated across the last 8 reversals. The average test duration was 1 hour.

Analysis

We estimated individual auditory filter shapes by fitting a rounded exponential (roex) function to the eight tone thresholds as a function of normalised frequency of the notch edge using a least squares routine [S18, S21] (for more details see [S19]). We computed individual predicted detection thresholds of each notch width and the pass-band (p) (i.e. slope) of the auditory filter.

Amplitude modulation test

Stimuli and presentation software

To assess abilities to perceive temporally modulated acoustical stimuli, we used an amplitude modulation test adapted from [S22]. The stimuli consisted of 2 s long speech-shaped noise (cf. for details on stimuli generation see 'speech-in-noise test') that was sinusoidal amplitude modulated (AM) with 50% modulation depth. The modulation rate roved by ±10% between trials. We estimated modulation thresholds at two modulation rates in separate tests: AM at a rate of 2Hz (slow AM test) and 40Hz (fast AM test). Slow temporal changes in speech are associated with suprasegmental information, such as prosody, while rapid temporal changes are associated with phonemic information [S23]. All stimuli were created digitally at a sampling rate of 44.1 kHz and 16 bit resolution and presented using Matlab (version 7.7, The MathWorks, Inc., MA, USA). Responses were recorded via a keyboard and stimuli were presented at a level comfortable to the individual participant.

Procedure

Each trial consisted of the sequential presentation of two AM-stimuli with different modulation frequency. In a two-interval two-alternative forced-choice paradigm, participants indicated which of the two stimuli contained the higher modulation rate. The slow and the fast AM test each included three runs. Each run ended after 16 reversals using a two-down one-up staircase procedure. We estimated the participants' thresholds (corresponding to 71% correct) from a single run by taking the mean differences in modulation rate at the last six reversals [S24]. For each modulation rate, we defined the participants' overall thresholds as the mean threshold from the three runs. To familiarise the participants with the task, they performed a practice run before the first test started. Feedback about response accuracy was provided after each trial. The complete test took approximately 45 minutes.

Musical-processing tests

Web-based musical skill test

The online version of the MBEA [S5] is accessible on http://www.brams.umontreal.ca/onlinetest/. The test included three sub-tests: 'off-beat', 'mistuned', 'out-of-key'. In the off-beat test participants were asked to detect differences in meter between otherwise identical melodies. In the mistuned test participants were asked to detect the presence of a single tone mistuned by half a semitone within and in the out-of-key test participants were asked to detect the presence of a single tone mistuned to be out of key with the rest of the melody. Both AS and SP conducted this test at home. The average test duration was 15 minutes.

Musical instrument recognition

Stimuli and presentation software

Stimulus material consisted of 20 western single-instrument sound samples [S25] and images of each instrument. Images were downloaded from free websites and were modified with Adobe Photoshop CS4 (version 11.0.2; Adobe Systems, Inc., San Jose, CA, US). All auditory stimuli were peak amplitude scaled using PRAAT [S7]. This test was implemented in Presentation software (Neurobehavioural Systems, Inc., CA, USA) and responses were recorded via a response button box.

Procedure

After hearing an instrumental sound, participants selected the target instrument out of four alternatives. The response array contained names and pictures of the target and three distractor instruments, one within-family instrument and two from different instrument families. To familiarise participants with the task, they first took part in a practice session. No feedback was given. The test took approximately 5 minutes.

Analysis procedure for all tests

We compared differences between the scores of AS and SP and their respective control groups on the web-based and the laboratory-based tests using a modified t-test for single case studies [S26]. Differences with a probability $p < .05$ were considered as statistically significant. For tests for which we had a directed hypothesis we performed one-tailed t-tests. For tests for which such a directed hypothesis was not possible (c-value in the famous voice-recognition test, performances on the control tests, acoustic voice-processing tests, musical-processing tests, the notched-noise, and the amplitude modulations tests, interaction between voice-learning, control tests, and the audiometry) we computed two-tailed t-tests.

Supplemental Discussion

AS and SP's performance is not predicted by task difficulty

The performance differences between SP and AS and their respective controls are unlikely to be due to a general performance deficit in challenging auditory tasks because the unfamiliar voice-learning tasks were not the most difficult auditory tasks. The mean scores of the control groups on the vocal-emotion test (controls: 80.18%) indicate that this test seems to be more difficult than the voice-face test (controls: 88.20%). Nevertheless, AS and SP performed within the controls' range on the vocal-emotion test.

AS and SP's performance on the AM modulation test

When looking at the two AM modulation tests separately, AS and SP had both normal performances overall (AS: $p = .11$, SP: $p = .18$, Table S6, Figure S1) and in one of the tests (AS: slow AM $p = .66$; SP: fast AM $p = .97$). However, AS seemed to have some difficulties with faster modulations ($p = .02$) and SP with slower modulations ($p = .03$). We do not believe that these differences in AM discrimination play a causal role in AS and SP's voice-recognition difficulties. First, the differences between AS and SP's AM-discrimination thresholds and those of their controls (AS: 2.87 SDs for 40Hz; SP: 2.5 SDs for 2Hz) are quite small in comparison to the differences we observed in the voice tests (AS: up to 8.73 SDs, SP: up to 9.34 SDs). Second, there was no correlation between AM-discrimination thresholds and voice recognition in the control groups of AS and SP ($r(11) = -.06$, $p = .85$). Third, we controlled for potential prosody-driven strategies in the voice-learning tests that rely on slow temporal changes. This was done by using sentences with different syntactical structure in the training and the test: five-word declaratives were used in the training, and the test consisted of five- as well as two-word declaratives and five-word interrogatives. If voice recognition performance relied on slow temporal prosodic information, we would have expected a decreased recognition performance in the test items that differed in syntactical structure from the training items. However, there was no main effect of sentence type in SP or his controls (Welch-test: $F(74.13) = 2.21$, $p = .117$).

Possible sampling bias in our screening approach

In the present report, we combined multiple methods to search for cases with developmental phonagnosia. The performance means in the web-based test were only slightly and not significantly higher than in the randomly ascertained lab-based control group (Table S1). Similarly, other online tests of face recognition and musical ability do not have inherently stronger sampling biases than laboratory tests [S5, S27]. It is however unclear whether such tests sample a representative set of the population. For example, our web-based participants were younger and more often female than would be expected in a representative sample of the general population. We also do not know whether the participants were representative in terms of for example educational level, since this was not assessed in the web-based

sample. Although we were able to collect a large set of data from the German-speaking population, we cannot exclude the possible influence of a sampling bias, which could have an effect in either direction. We speculate that the true prevalence of phonagnosia might be higher than 2‰, since the return rate of our four-stage screening approach was relatively low, especially for the questionnaire (return rate = 23%). We assume that the low return rate can be attributed to either a lack of time or interest, or a reservation to report on a deficit (either voice-recognition deficit or other medical deficits that were asked for in the questionnaire).

Supplemental References

S1. Zimmermann, P., and Fimm, B. (2009). Testbatterie zur Aufmerksamkeitsprüfung: TAP, Version 2.2, (Psytest).

S2. Schmidt, K.H., and Metzler, P. (1992). Wortschatztest (WST). Hogrefe, Göttingen.

S3. Härting, C., Markowitsch, H.J., Neufeld, H., Calabrese, P., Deisinger, K., and Kessler, J. (2000). Wechsler Gedächtnis Test-Revidierte Fassung (WMS-R): Deutsche Adaptation der revidierten Fassung der Wechsler Memory Scale. Göttingen: Hogrefe.

S4. Schuri, U., and Benz, R. (2000). Gesichter-Namen-Lerntest, (Swets Test Services).

S5. Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L.L., Gagnon, B., Trimmer, C.G., Paquette, S., and Bouchard, B. (2008). On-line identification of congenital amusia. Music Percept *25*, 331-343.

S6. Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia *9*, 97-113.

S7. Boersma, P., and Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.14).

S8. Association, A.S.-L.-H. (2005). Guidelines for manual pure-tone threshold audiometry.

S9. Duchaine, B., and Nakayama, K. (2006). The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia *44*, 576-585.

S10. Macmillan, N.A., and Creelman, C.D. (2004). Detection theory: A user's guide, (Psychology press).

S11. Brungart, D.S., Simpson, B.D., Ericson, M.A., and Scott, K.R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. Journal of the Acoustical Society of America *110*, 2527-2538.

S12. Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. Perception & psychophysics *49*, 227-229.

S13. Wendt, B. (2007). Analysen emotionaler Prosodie, (P. Lang).

S14. Oatley, K., and Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. Cognition and emotion *1*, 29-50.

S15. Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. pp. 3933-3936.

S16. Smith, D.R.R., Patterson, R.D., Turner, R., Kawahara, H., and Irino, T. (2005). The processing and perception of size information in speech sounds. J.Acoust.Soc.Am. *117*, 305-318.

S17. Fitch, W.T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J Acoust Soc Am *106*, 1511-1522.

S18. Patterson, R.D., Nimmo-Smith, I., Weber, D.L., and Milroy, R. (1982). The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. J Acoust Soc Am *72*, 1788-1803.

S19. Herrmann, B., Henry, M.J., Scharinger, M., and Obleser, J. (2013). Auditory filter width affects response magnitude but not frequency specificity in auditory cortex. Hearing research *304*, 128-136.

S20. Leek, M.R. (2001). Adaptive procedures in psychophysical research. Percept Psychophys *63*, 1279-1292.

S21. Moore, B.C. (2005). Basic psychophysics of human spectral processing. International review of neurobiology *70*, 49-86.

S22. Teki, S., Kumar, S., von Kriegstein, K., Stewart, L., Lyness, C.R., Moore, B.C., Capleton, B., and Griffiths, T.D. (2012). Navigating the auditory scene: an expert role for the hippocampus. J Neurosci *32*, 12251-12257.

S23. Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci *336*, 367-373.

S24. Levitt, H. (1971). Transformed up-down methods in psychoacoustics. J Acoust Soc Am *49*, Suppl 2:467+.

S25. Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J.R., Schweinberger, S.R., Warren, J.D., and Duchaine, B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia *47*, 123-131.

S26. Crawford, J.R., and Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. The Clinical Neuropsychologist *12*, 482-486.

S27. Germine, L., Nakayama, K., Duchaine, B.C., Chabris, C.F., Chatterjee, G., and Wilmer, J.B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev *19*, 847-857.