ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Statistical learning of speech, not music, in congenital amusia

Isabelle Peretz,[1] Jenny Saffran,[1,2] Daniele Schön,[3] and Nathalie Gosselin[1]

[1]International Laboratory of Brain, Music and Sound Research (BRAMS), Department of Psychology, University of Montreal, Montreal, Canada. [2]University of Wisconsin, Madison, Wisconsin. [3]Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

Address for correspondence: Isabelle Peretz, BRAMS-Pavillon 1420 Mont-Royal, Université de Montréal, Montreal, Quebec, Canada H2V 4P3. Isabelle.Peretz@umontreal.ca

The acquisition of both speech and music uses general principles: learners extract statistical regularities present in the environment. Yet, individuals who suffer from congenital amusia (commonly called tone-deafness) have experienced lifelong difficulties in acquiring basic musical skills, while their language abilities appear essentially intact. One possible account for this dissociation between music and speech is that amusics lack normal experience with music. If given appropriate exposure, amusics might be able to acquire basic musical abilities. To test this possibility, a group of 11 adults with congenital amusia, and their matched controls, were exposed to a continuous stream of syllables or tones for 21-minute. Their task was to try to identify three-syllable nonsense words or three-tone motifs having an identical statistical structure. The results of five experiments show that amusics can learn novel words as easily as controls, whereas they systematically fail on musical materials. Thus, inappropriate musical exposure cannot fully account for the musical disorder. Implications of the results for the domain specificity of statistical learning are discussed.

Keywords: statistical learning; speech; music; pitch disorder; amusia

Introduction

Statistical learning is the process by which learners rapidly acquire structured information from variable environmental inputs in the absence of explicit feedback.[1] This mechanism operates in numerous domains, such as speech, music, and visual geometric figures. This powerful mechanism is considered to be fundamental to the acquisition of language because statistical learning plays a critical role in speech segmentation.[2,3] Several studies demonstrate comparable learning with tones,[4] musical timbres,[5] and sung syllables.[6] Because the type of computation appears to operate equivalently across syllables and tones, statistical learning is conceived as a domain-general mechanism that encompasses the acquisition of both music and language.

This raises the question of why such a domain-general mechanism does not function properly in the case of congenital amusia. Amusic individuals

typically display normal language abilities and poor musical skills. For example, they can normally recognize the lyrics of songs, although they fail to recognize the melody component of the same song.[7] One possibility is that amusics have had impoverished learning experiences with music because their brain was not "ready" for or tuned to music at the appropriate time during development. This delay may have been exacerbated by a lifelong history of musical failures and lack of interest for music. Indeed, Conway *et al.*[8] show that a period of auditory deprivation may impair general sequencing abilities, which in turn may explain why some deaf children still struggle with language following cochlear implantation. The goal of the present study was to test whether amusics might be able to acquire basic musical abilities by way of statistical learning if given appropriate exposure.

To address this aim, a group of 11 adults with congenital amusia and their matched controls were

exposed to a continuous stream of syllables or tones that were organized according to a set of simple statistical regularities (e.g., the syllable /pa/ tends to be followed by the syllable /bi/, or the tone A4 tends to be followed by F#4). After this familiarization phase, the participants were asked to judge the familiarity of three-syllable nonsense words or three-tone motifs, defined by the same transitional probabilities. As mentioned earlier, normal adults rapidly detect the regularities that link together the syllables by discriminating familiar groups of syllables (or tones) better than syllables (tones) that span the group boundary. This effect was first discovered by Saffran *et al.*[2,4] Thus, as a starting point, we used Saffran's material for adults in Experiment 1 (syllables) and in Experiment 2 (pure tones). Because amusics failed to learn the pure tone structure, we tested them further with a reduced set of pitches (Experiment 3) and of targets (Experiment 4). Finally, we tested the learning of pitch structure indirectly by presenting sung syllables in Experiment 5.

General method

Participants

Eleven amusic adults and 13 matched controls (who had no musical education and no musical impairment) were tested (see Table 1 for their background information). They were considered amusic (or not) on the basis of their scores on the Montreal Battery of Evaluation of Amusia (MBEA).[9] The battery involves six tests (180 trials) that assess various music processing components. Three tests assess the ability to discriminate pitch changes in melodies (melodic tests), and three tests assess the ability to

discriminate rhythmic changes, meter, and memory, respectively. Each amusic participant obtained a global score and a melodic score that were both two standard deviations below the control participants (Table 1).

Procedure

All subjects were francophone and tested individually. They were instructed to listen to a nonsense language. They were further told that the language contained words (or tone motifs), but no meaning or grammar. Their task was to figure out where the words (motifs) began and ended. Note that we could not deceive the subjects as we used a within-subjects design and, thus, after having been tested with syllables (Experiment 1), they were also later tested with tones, although in different sessions. After the familiarization phase, subjects were presented with a forced-choice test. They were instructed to indicate which of the two sequences sounded more like a word (motif) from the familiarized language. To do so, they pressed either "1" or "2" on the keyboard of the computer, corresponding to whether the familiar sequence was played first or second in that trial. The two test sequences were separated by 750 ms, with an intertrial of five seconds. Stimuli were presented over speakers at a comfortable level.

Experiment 1: syllables

The first step was to determine whether amusics can learn novel words determined by adjacent probabilities. To this aim, we used the material and procedure of Ref. 3. It consists of four consonants (p, t, b, d) and three vowels (a, i, u) combined into 12 CV syllables making up six words (babupu, bupada,

Table 1. Mean age, education, and gender for amusic and control participants; mean percentages of correct response (standard deviations) obtained on the MBEA (global score) for the 11 amusics and the 13 matched controls. Chance level is 50% in all tests. Group differences were assessed by way of bilateral *t* tests.

Participants	Amusics	Controls	*t* tests
Gender	4M, 7F	5M, 8F	
Age (SD)	59.1 (9.5)	58.9 (6.9)	$t(22) = 0.05$, *n.s.*
Years of education (SD)	16.9 (1.6)	17.2 (2.5)	$t(22) = -0.28$, *n.s.*
Musical experience level[a] (SD)	1.9 (1.0)	1.9 (0.8)	$t(21) = -0.02$, *n.s.*
MBEA% global score (SD)	64.5 (7.6)	88.0 (3.7)	$t(22) = -9.85$, $P < 0.001$

[a]level 1 = less than 1 year; 2 = between 1 and 3 years; 3 = between 4 and 6 years; 4 = between 7 and 10 years; 5 = more than 10 years.
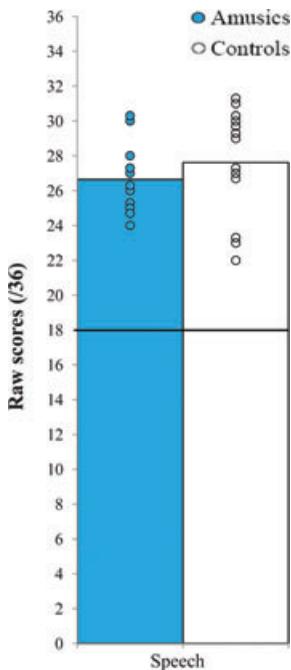
Figure 1. Performance by amusic and control participants in Experiment 1. Circles represent the number correct (out of a possible 36) for individual subjects; columns represent the mean; chance = 18.

dutaba, patubi pidabu, and tutibu). Note that some syllables occurred in more words than others; for example, "bu" occurred in four words, while "ta" occurred in only one. Thus, the transitional probabilities between syllables within words varied (range: 0.3–1.0) and were higher than the transitional probabilities between syllables spanning a word boundary (range: 0.1–0.2). The words were randomized and combined into a 21-minute stream of continuous speech with no pauses between words. The test consisted of 36 items. Each item paired a word from the stream with a part-word sequence, which consisted of a sequence that crossed a word boundary (e.g., dutaba–bapatu).

Results are presented in Figure 1. The individual scores obtained by both amusics and controls were significantly different from chance, where chance equals 18 (50%; t [10] = 14.23 and t [12] = 11.01 for amusics and controls, respectively, both $P < 0.001$ by two-tailed tests). Moreover, amusics (74.0% correct) performed as well as controls (76.7%; t[22] = 0.887). The results are very similar to those (76%) obtained in university students in Ref. 3.

Thus, amusics can learn novel words on the basis of transitional probabilities as well as controls.

Experiment 2: pure tones

Here we tested whether the same amusic participants can track transitional probabilities in a tone sequence analogous in structure to the speech materials from Experiment 1. To do this, we used the material and procedure of Saffran *et al.*,[4] who substituted a distinct tone for each of the 12 syllables from which the novel words were created in Experiment 1 (e.g., BU became the musical note D). The 12 tones were pure tones in the same octave, starting at middle C within a chromatic set, with a duration of 330 milliseconds. Each of the six trisyllabic nonsense words (e.g., BUPADA) from the novel speech stream was thereby translated into a sequence of three musical tones (e.g., DFE). Six three-tone motifs (ADB, DFE, GG#A, FCF#, D#ED, and CC#D) were then concatenated together, in a random order and with no silence between motifs, to generate a continuous stream identical in statistical structure to the speech stream tested in Experiment 1. The tones stream was presented for 21 min and followed by 36 test trials. Each trial consisted of a pair of three-tone motifs, one of which was a tone motif from the familiarization stream, and the other one straddled a motif boundary.

The pure tone motifs were impossible to learn for 6 out of 7 controls as well as for 7 out of 10 amusics. None of the groups performed above chance level (60%; t [6] = 0.00; and 48%; t [9] = 0.25, for controls and amusics, respectively, both $P_s > 0.05$). This was unexpected given that Saffran *et al.*[4] found successful learning with similar tone stimuli with college-aged participants. However, Evans, Saffran and Robe-Torres[10] also found that these tone sequences were particularly difficult for children with specific language impairment to learn. Indeed, the children performed above chance after 40 min of exposure to a similar speech stream, while failing to acquire the tone stream after the same amount of exposure. Thus, to make the tones easier to learn, we changed the material in Experiments 3 and 4 while keeping the same procedure.

Experiment 3: diatonic piano tones

In order to simplify the task at least for the musically unimpaired controls, we reduced the 11 chromatic pitches to 8 diatonic tones. All pitches were
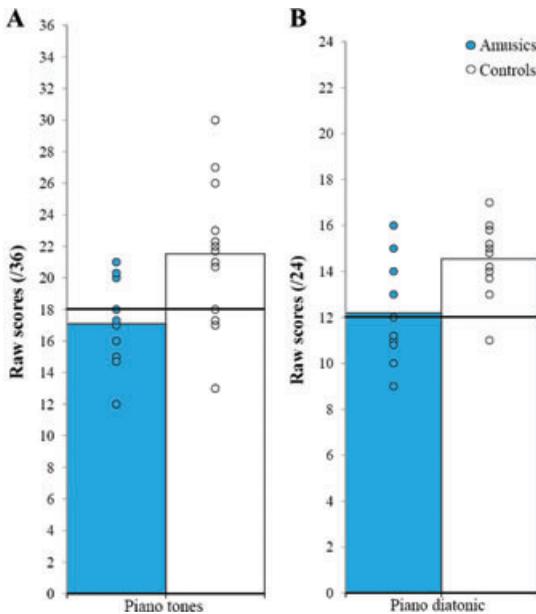
Figure 2. (A) Performance by amusic and control participants in Experiment 3. Circles represent the number correct (out of a possible 36) for individual subjects; columns represent the mean; chance is 18. (B) Performance by amusic and control participants in Experiment 4. Circles represent the number correct (out of a possible 24) for individual subjects; columns represent the mean; chance = 12.

diatonic and taken from the octave above middle C. The six motifs were EDC, C*FG, CEF, GC*D, DGC, and FED (* = one octave above middle C). Furthermore, to facilitate the forced choice between the motifs and the foils, the transitional probabilities between the individual pitches in the test foils were low (CGF, ECC*, FDF, DFC*, CDE, and C*GD). That is, adjacent pitches in a foil did not occur in the tone motifs. For example, in the test foil CGF, learners did not hear CG or GF in the target motifs (as done in Ref. 11, to be discussed later). As previously, the foil pitches spanned motif boundaries. Finally, instead of using pure tones, the pitches were presented with the timbre of a piano. Otherwise, the task was identical. Participants were presented with the six motifs for 21 min after which they had to select the familiar motif in 36 test trials.

The results are presented in Figure 2A. This time, controls succeeded to perform above chance as a group with 59.8% ($t[12] = 2.78$, $P < 0.05$). None of the 11 amusics performed above chance (with 47.5%; $t[9] = 1.03$) and performed significantly below controls ($t[21] = 2.694$, $P < 0.02$). Thus, amu-

sics failed to learn the novel motifs whereas most of their matched controls learned them. However, four out of the 12 controls performed at chance despite the fact that the diatonic set allowed them to use tonal encoding of pitch as a possible memory code.

Experiment 4: four motifs

In order to boost performance so that all musically unimpaired controls could achieve recognition of the novel tone motifs above chance, we reduced the number of targets to recognize in the test trials. Familiarization consisted of the same material as in Experiment 3 and lasted 21 minutes. Only the test pairs were changed. The number of motifs to recognize in the test was decreased from six to four. These four motifs (EDC, C*FG, DGC, and CEF) were selected so as to be as distinct as possible in terms of pitch contour. The foils were the reverse orderings of the motifs so as to keep the same intervals; the transitional probabilities between tones in the foils were always zero. The test phase consisted of 24 trials instead of 36.

As can be seen in Figure 2B, the changes were successful in bringing the scores of all but one matched control above chance ($t[12] = 2.78$, $P < 0.05$). Amusics' scores remained at chance ($t[9] = 1.03$) and below controls ($t[21] = 2.69$, $P < 0.05$). Among the four amusics who obtained scores above chance here, only two obtained similar scores in Experiment 3. Note that we also tried 40 min of exposure with a few amusics, with no more success.

Experiment 5: syllables and sung tones

Amusics repeatedly failed to learn novel tone motifs, although successfully learned novel words organized along the same statistical principles. This result indicates that lack of adequate musical exposure is probably not the critical factor, at least within the limited time window of exposure used here. Rather, the results suggest that statistical learning of speech does not involve the same codes or processes as statistical learning of music.

However, it is generally easier to extract statistical rules from speech than music (see Ref. 12 for a review). Because tones are more difficult to learn than syllables, the same statistical mechanism may operate at different levels of efficiency in both controls and amusics; for one reason or another, amusics
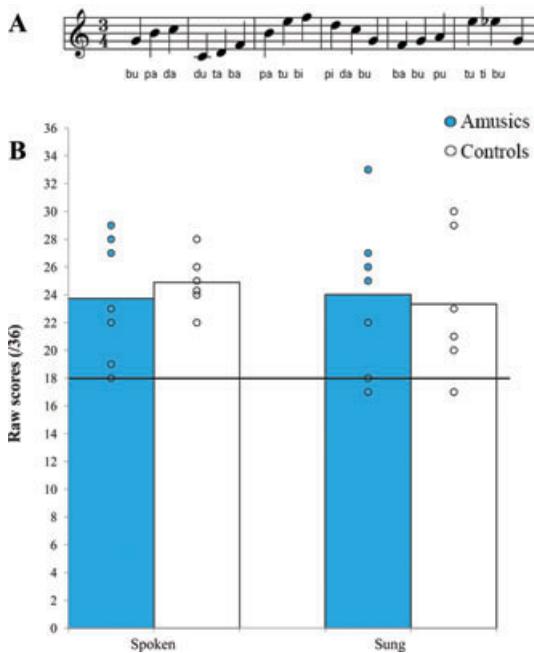
Figure 3. (A) Representation of words and motifs used as targets in Experiment 5; (B) Performance by amusic and control participants after familiarization with spoken or sung materials in Experiment 5. Circles represent the number correct (out of a possible 36) for individual subjects; columns represent the mean; chance = 18.

would not fully engage in the tone motifs processing.

One solution to avoid this efficiency difference is to test novel word recognition and examine whether the addition of melodic structure improves learning. This has been previously studied by Schön *et al.*,[6] who showed that novel words were better learned when sung, provided that the motif structure coincided with the word boundaries, both defined in terms of transitional probabilities.

Similarly, here amusic and control participants were presented with 21 min of either spoken syllables (as in Experiment 1) or sung tones. The grouping of the spoken syllables and sung syllables were determined by transitional probabilities. Although the familiarization phase consisted of either spoken words or sung words, it was followed by the same test phase. In the latter, both words and foils were presented in a spoken format (not sung; Fig. 3A).

The advantage of this procedure is that the learning of the motif structure is assessed indirectly. This may provide a better test of amusics' ability to ex-

tract pitch structure as they can show evidence of statistical learning without awareness.[13]

Method
The structure of the continuous stream of spoken syllables was identical to that used in Experiment 1. The stream was generated using a male French voice setting on the Mbrola speech synthesizer (http://tcts.fpms.ac.be/synthesis/mbrola.html). The result was a monotone and continuous stream of syllables. The foils consisted of either the last syllable of a word plus the first syllable pair of another word, or the last syllable pair of a word plus the first syllable of another word. For instance, the last syllable of the word "dutaba" was combined with the first two syllables of "patubi" to create the partial word foil "bapatu."

In the sung version, each of the 11 syllables was associated with a distinct tone (Fig. 3A). The eleven tones were C5, D5, F5, G5, A5, B5, C6, Db6, D6, E6, and F6. Thus, the statistical structure of the spoken and sung sequence was identical and superimposed in the sung version, with constant syllable–pitch mapping. The sung sequence was also synthesized using Mbrola.

Seven amusics and seven controls were tested with the spoken and sung version. The order of presentation of the spoken material and sung material was counterbalanced across participants.

Results and comments
Like Experiment 1, amusics performed above chance in the spoken version of the material (66%; $t[6] = 3.45$ $P < 0.05$). They also performed above chance with the sung version (67%; $t[6] = 2.87$, $P < 0.05$; Fig. 3B). This provides converging evidence that amusics can learn novel words on the basis of statistical learning. Their scores did not differ from the scores obtained by their matched controls who also performed above chance in both the spoken condition (68%; $t[6] = 3.45$, $P < 0.005$) and in the sung condition (65%; $t[5] = 2.53$, $P = 0.053$).

An ANOVA including conditions (spoken, sung) as within-subjects factor and groups (amusics, controls) as between-subjects factor revealed no significant effects for condition ($F[1, 11] = 0.463$, $\eta^2 = 0.04$) or group ($F < 1$), nor an interaction between these two factors ($F < 1$). Thus, we failed to replicate the advantage of the sung over the spoken words obtained by Schön *et al.*[6]

This failure might be because of two differences in the experimental design. The first is that the duration of the familiarization phase was three times longer in the current study than in Schön *et al.*'s study. Thus, participants may reach a plateau with both materials. The second difference is that while Schön *et al.*[6] used a between-subjects design; we used a within-subjects design. Some transfer of learning may have taken place. Marcus *et al.*[14] have shown that infants are better able to extract rules from sequences of musical tones if they first hear those rules instantiated in sequences of speech. We did not find support for such an effect here. The performance on the sung words of the subjects who started with the spoken version was 65.9% and did not differ from the 68.9% obtained by the subjects who started with the sung version ($t[10] = 0.35$, $P = 0.73$). Note that unlike previous studies conducted with students, we tested small samples of subjects. We computed that we would need to test 36 amusics to show an advantage for the sung material over the spoken one. This is simply not realistic.

Like Schön *et al.*,[6] we observed that words with higher transitional probabilities tended to be better recognized than words with lower transitional probabilities (70.5% and 62.8%, respectively), across conditions and groups ($F[1, 12] = 3.74$, $P = 0.077$). Amusics did not differ from controls ($F[1, 11] = 0.01$, $P = 0.918$) and the mode of familiarization (sung versus spoken) did not influence learning ($F[1, 11] = 0.21$, $P = 0.653$).

General discussion and conclusions

Individuals with congenital amusia can learn novel words but fail to learn novel pitch motifs organized according to the same statistical properties. These results suggest that statistical learning is not mediated by a single processing system but by two systems specialized for processing syllables and tones, respectively. It is likely that the speech and musical input codes adjust statistical learning to their processing needs. In other words, the input and output of the statistical computation may be domain-specific while the learning mechanism is not.[15,16]

The present findings also suggest that amusics will be unable to develop a normal capacity for music, although they can acquire a novel language. This dissociation was clearly present in the data despite the fact that amusics had equal opportunities

to learn the two materials. However, amusics may show evidence of statistical learning for pitch intervals[11] without awareness.[13] Thus, the experimental setting used here may not be adequate to reveal implicit learning in small samples of amusic adults. Alternatively, the musical representations acquired implicitly may not be robust enough to sustain the interference created by the different test trials.

Acknowledgments

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Aslin, R.N. & E.L. Newport. 2009. What statistical learning can and can't tell us about language acquisition. In *Infant Pathways to Language: Methods, Models, and Research Disorders*. J. Colombo, P. McCardle & L. Freund, Eds.: 15–29. Erlbaum. New York.

2. Saffran, J.R., R.N. Aslin & E.L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* **274:** 1926–1928.

3. Saffran, J.R., E.L. Newport & R.N. Aslin. 1996. Word segmentation: the role of distributional cues. *J. Mem. Lang.* **35:** 606–621.

4. Saffran, J.R. *et al.* 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* **70:** 27–52.

5. Tillmann, B. & S. McAdams. 2004. Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn.* **30:** 1131–1142.

6. Schon, D. *et al.* 2008. Songs as an aid for language acquisition. *Cognition* **106:** 975–983.

7. Ayotte, J., I. Peretz & K.L. Hyde. 2002. Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* **125:** 238–251.

8. Conway, C.M. *et al.* 2011. Implicit sequence learning in deaf children with cochlear implants. *Dev. Sci.* **14:** 69–82.

9. Peretz, I., A.S. Champod & K.L. Hyde. 2003. Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Ann. N.Y. Acad. Sci.* **999:** 58–75.

10. Evans, J., J.R. Saffran & K. Robe-Torres. 2009. Statistical learning in children with Specific Language Impairments. *J. Speech Lang. Hear. Res.* **52:** 321–335.

11. Omigie, D. & L. Stewart. 2011. Preserved statistical learning of tonal and linguistic material in congenital amusia. *Front. Psychol.* **2:** 109.

12. Francois, C. & D. Schon. 2011. Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex.* **21:** 2357–2365.

13. Peretz, I. *et al*. 2009. The amusic brain: in tune, out of key, and unaware. *Brain* **132:** 1277–1286.

14. Marcus, G.F., K.J. Fernandes & S.P. Johnson. 2007. Infant rule learning facilitated by speech. *Psychol. Sci.* **18:** 387–391.

15. Peretz, I. 2006. The nature of music from a biological perspective. *Cognition* **100:** 1–32.

16. Saffran, J.R. & E.D. Thiessen. 2006. Domain-general learning capacities. In *Handbook of Language Development*. E. Hoff & M. Shatz, Eds.: 68–86. Blackwell. Cambridge.