



HAL
open science

Simple frequency ratios naturally make precisely perceived melodies

Laurent Demany, Catherine Semal, Daniel Pressnitzer

► **To cite this version:**

Laurent Demany, Catherine Semal, Daniel Pressnitzer. Simple frequency ratios naturally make precisely perceived melodies. 2024. hal-04801514

HAL Id: hal-04801514

<https://hal.science/hal-04801514v1>

Preprint submitted on 25 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Simple frequency ratios naturally make precisely perceived melodies

Laurent Demany,¹ Catherine Semal,¹ and Daniel Pressnitzer²

¹Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, CNRS, Université de Bordeaux, 33076 Bordeaux, France

²Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, CNRS, PSL University, 75005 Paris, France

Summary

Almost all human music is built on discrete scales of pitch [1]. Culturally prominent scales, such as the diatonic major scale of Western music, make use of the simple frequency ratios 2:1, 3:2, and 4:3 between notes [2]. It is generally believed that these ratios were chosen to optimize the consonance of simultaneous notes [3-7]. Alternatively, or in addition, it is conceivable that these ratios are intrinsically advantageous for the perceptual encoding of melodies [8-12]. Here we provide behavioral support for this hypothesis. In three experiments, young Western adults had to detect pitch anomalies ("sour notes") in partly random pure-tone melodies based on various musical scales, including novel ones. The task did not require any musical knowledge. Most importantly, the listeners were extensively trained in order to saturate familiarity with the scales: for a given scale and listener, more than 2000 (up to 5280) trials were run. Practice largely improved performance. This occurred even for the diatonic major scale, suggesting that performance in our task was not modulated by previous musical enculturation [13-14]. Frequency ratio simplicity also favored performance. Crucially, its benefit was not smaller in the final test sessions, when performance for each scale was presumably optimal and no longer improvable by practice, than in the initial test sessions. Thus, frequency ratio simplicity appeared to be intrinsically advantageous, rather than advantageous merely due to familiarity. The naturalness of melodic intervals defined by simple frequency ratios is likely to have contributed to the cultural selection of musical scales.

Results

On each trial of three experiments performed on young Western adults, the participant was presented with two successive melodies of pure tones (see Figure 1A). Melody 1 consisted of a random ordering of the notes (degrees) defining a musical scale, in a frequency range which was also randomly selected. Melody 2 was another random ordering of the same tones, either without any error or with a 1-semitone shift in one tone, corresponding to a variable degree of the scale; this tone then became out of scale. The participant had to indicate if melody 2 contained a shifted tone ("sour note") or not. It could be expected that, with practice, performance would improve in consequence of perceptual learning. Our main question was about a possible advantage of simple frequency ratios after the saturation of perceptual learning.

The rationale, illustrated in Figure 1B, was as follows. Consider two scales, X and Y, with a superiority of X over Y regarding frequency ratio simplicity. At the onset of practice, performance is expected to be better for X than for Y, either because frequency ratio simplicity is naturally propitious to melody perception or because Y is less familiar than X. Practice will increase familiarity, if it can be increased. Suppose that at the end of practice, performance for each scale has reached a plateau and can no longer be improved. This implies that the positive effect of familiarity on performance is as large as it can be. Therefore, if the initial advantage of X was entirely due to familiarity, this advantage should have disappeared at the end of practice. If, on the contrary, the advantage was entirely due to the intrinsic properties of the scales, then it could be equally large at the end of practice.

Experiment 1

In experiment 1, we employed four scales labeled as *Major_N*, *Major_S*, *WT_N*, and *WT_S* (see Figure 2A and Methods). *Major_N* was the equally tempered major scale prominent in Western music; it spanned one octave – the musical interval defined by a frequency ratio of 2:1 – and also made use of frequency ratios nearly equal to 3:2 and 4:3. *Major_S* was obtained by a 20% stretching of *Major_N*; this stretching excluded any ratio close to 2:1 or 3:2. *WT_N* is known as the "whole-tone scale", dividing the octave into six equal intervals; except for the octave itself, it is devoid of simple frequency ratios. *WT_S* was obtained by a 20% stretching of *WT_N*; this stretching preserved the 2:1 ratio and produced ratios close to 3:2 and 4:3. If melody perception takes advantage of frequency ratio simplicity, one could

expect that, in our task, scale stretching would be deleterious from *Major_N* to *Major_S*, but advantageous from *WT_N* to *WT_S*.

In Phase 1 of the experiment, 32 test sessions (consisting each of 220 trials; see Methods) were run: 16 sessions devoted to *Major_N* interleaved with 16 sessions devoted to *WT_N*. For both scales (see Figure 2B), performance initially improved and then stabilized after about eight sessions, i.e., half of the phase. Performance was better for *Major_N* than for *WT_N* and this difference remained essentially constant throughout the phase. A repeated-measures ANOVA [Scale \times Session] confirmed that Scale [$F(1, 14) = 17.70, p < 0.001, \eta_p^2 = 0.56$] and Session [$F(15, 210) = 25.11, p < 10^{-6}, \eta_p^2 = 0.64$] had significant main effects while their interaction was not significant [$F(15, 210) = 0.82, p = 0.66, \eta_p^2 = 0.05$]. An ANOVA restricted to sessions 1-8 for each scale revealed again a significant effect of Session [$F(7, 98) = 12.84, p = 0.003, \eta_p^2 = 0.67$], while this was not the case for an ANOVA restricted to sessions 8-16 [$F(8, 112) = 0.79, p = 0.61, \eta_p^2 = 0.05$]. The fact that performance improved to the same extent for *Major_N*, which was familiar to all participants at the onset of the experiment, and for *WT_N*, which was certainly much less familiar, is remarkable: it suggests that musical enculturation had essentially no effect on performance in our task.

In Phase 2, *Major_N* and *WT_N* were replaced by *Major_S* and *WT_S*, for 20 test sessions (10 for each scale). As shown by Figure 2B, performance was very poor for *Major_S* and much better for *WT_S* [$F(1, 14) = 130.32, p < 10^{-6}, \eta_p^2 = 0.90$]. A significant effect of Session was again found, for both scales [$F(9, 126) \geq 2.92, p \leq 0.004$]. However, whereas practice improved performance markedly in the case of *WT_S*, performance for *Major_S* was still quite poor at the end of this phase. Given that performance had stabilized at the end of Phase 1, the pronounced improvement observed for *WT_S* indicates that the benefit of practice with a given scale was at least partly specific of this scale.

Phase 2 was followed by a short Phase 3, replicating Phase 1 in 8 test sessions intended to confirm that optimal performance for *Major_N* and *WT_N* had been reached in Phase 1. This was indeed confirmed as performance was not better in Phase 3 than at the end of Phase 1 (see Figure 2B).

It seems reasonable to assume, more generally, that practice allowed the participants to reach at least approximately their best possible performance for each of the four scales. Under this assumption, differences between the scales with respect to final performance can be legitimately considered as due mainly to intrinsic properties of the scales rather than to familiarity. We assessed final performance by the mean d' obtained in the last three sessions of Phases 1 and 3 for *Major_N* and *WT_N*, and the last three sessions of Phase 2 for *Major_S*

and *WT_S*. The results are displayed in Table 1. There was no significant difference between *Major_N* and *WT_S* [$t(15) = 1.73$, $p = 0.104$], for which performance was best. Both of these scales differed significantly from *WT_N*, for which performance was lower [$t(15) \geq 3.85$, $p \leq 0.0016$], but significantly better than performance for *Major_S* [$t(15) = 10.18$, $p < 10^{-5}$]. The high final performance obtained for *WT_S*, in spite of the initial unfamiliarity of this scale, suggests again that musical enculturation was unimportant in the experiment.

The four scales differed from each other with respect to two crossed factors: a factor of "structure", differentiating *Major_N* and *Major_S* from *WT_N* and *WT_S*, and a factor of frequency ratio simplicity giving a potential advantage to *Major_N* over *Major_S* and to *WT_S* over *WT_N*. From previous research [15-16], it could be expected that structure would have a systematic effect, giving an advantage to the two "Major" scales because they were asymmetric whereas the two "WT" scales were symmetric. But this was not the case: structure had no systematic effect. In contrast, frequency ratio simplicity did have a systematic effect. Clearly, this intrinsic feature of the scales was the chief modulator of final performance.

Experiment 2

The very poor performance obtained for *Major_S* in experiment 1 indicates that participants were particularly sensitive to the presence or absence of the frequency ratio 2:1, one octave. In experiment 2, we focused on the frequency ratio 3:2, known as the "perfect fifth". Only two scales were used: those depicted in the upper half of Figure 3A. They were actually two modes of the equally tempered diatonic scale: the *Phrygian* mode and the *Locrian* mode. They differed from each other with respect to a single degree, the fifth one. In the *Phrygian* scale, this degree is positioned 7 semitones (a frequency ratio of $2^{7/12}$, very close to 3:2) above the first degree. In the *Locrian* scale, the corresponding interval is 6 semitones, a ratio equal to $\sqrt{2}$. The experimental melodies were constructed in the same manner as in experiment 1, except for the choice of the shifting tone when a tone was shifted from melody 1 to melody 2 (see Figure 3A and Methods). Twenty test sessions were run for each participant: 10 *Phrygian* sessions interleaved with 10 *Locrian* sessions.

Figure 3B shows how performance depended on the Scale and Session factors. A repeated-measures ANOVA found a highly significant effect of Scale [$F(1, 15) = 30.06$, $p < 10^{-4}$, $\eta_p^2 = 0.67$] as well as Session [$F(9, 135) = 14.78$, $p < 10^{-6}$, $\eta_p^2 = 0.50$], and a marginally significant interaction [$F(9, 135) = 2.31$, $p = 0.057$, $\eta_p^2 = 0.13$]. Performance was better for the *Phrygian* scale. Practice improved performance for both scales, but Figure 3B shows that the improvement was larger in the *Phrygian* condition. The latter trend was found to be

significant by a paired t-test in which the dependent variable was the slope of regression lines fitted to the individual data obtained from session 1 to session 10 [$t(15) = 2.89, p = 0.011$].

These results definitely support the idea that the musical interval known as the perfect fifth is a natural melodic interval. If, in the present experiment, the initial advantage observed for the *Phrygian* scale had been due to lesser familiarity of the *Locrian* scale, then the advantage should have decreased with practice. We found just the opposite, indicating that, in our task, frequency ratio simplicity can facilitate perceptual learning.

Experiment 3

Experiment 3 dealt with the frequency ratio 4:3, known as the "perfect fourth" and very close to $2^{5/12}$ (5 semitones). The two scales used in this experiment were labeled as *Phry* and *Loc*, because they were respectively parts of the *Phrygian* and *Locrian* scales (see Figure 3A). They differed from each other only with respect to their fourth degree, 5 semitones above the first degree in *Phry* but 6 semitones above it in *Loc*. The experiment consisted of 12 *Phry* sessions interleaved with 12 *Loc* sessions. For four of the 16 participants, a second run identical to the first one was performed, after a mean delay of 50 days (range: 22-74 days).

Figure 3C shows the results of the initial run (16 participants). There was a significant effect of Scale [$F(1, 15) = 32.27, p < 10^{-4}, \eta_p^2 = 0.68$] as well as Session [$F(11, 165) = 16.31, p < 10^{-6}, \eta_p^2 = 0.50$], but no significant interaction [$F(11, 165) = 1.06, p = 0.39, \eta_p^2 = 0.07$]. Performance was definitely better for *Phry* than for *Loc*. For each scale, d' significantly varied from session 1 to session 6 [$F(5, 75) \geq 14.31, p < 10^{-6}, \eta_p^2 \geq 0.49$], but not from session 6 to session 12 [$F(6, 90) \leq 1.40, p \geq 0.22, \eta_p^2 \leq 0.085$].

The four participants tested in a second run were selected mainly because, in the initial run, they showed particularly protracted performance improvements. Their complete results are displayed in Figure 3D. During the second run, there was no significant improvement. For each scale, indeed, performance remained somewhat worse than at the end of the initial run. This may largely be a statistical artefact resulting from the biased selection of these participants. Most importantly, performance remained better for *Phry* than for *Loc*. This was true for each participant in each of the three final sessions, after 4620 trials with each scale.

Overall, the experiment clearly supports the hypothesis that the perfect fourth is a natural melodic interval.

Discussion

Human melody perception is undoubtedly dependent on both nature and nurture. In the present study, nurture was clearly at work since practice generally improved performance to a large extent. Experiment 1 indicated that the improvement reflected, at least in part, scale-specific perceptual learning. However, in view of our results, the power of nurture appears to be limited since, for instance, performance for *Major_S* was still very poor after 10 test sessions. This finding is in stark contrast with previous suggestions that adult Westerners are able to learn quite rapidly (less than one hour) to perceive accurately melodies based on an initially unknown musical scale poor in simple frequency ratios [16-19].

Effects of nature on melody perception have also been previously identified, in particular via cross-cultural comparisons [20]. What we add here is that nature makes some melodic intervals perceptually special, even between pure tones. Past experiments with infants led to the same suggestion [8-12]. However, these experiments have been considered as not fully convincing [21] because musical enculturation may start very early in life [22]. Our study brings novel evidence from adults, in a task where enculturation was found to be irrelevant. Both sets of results are at odds with the widely-held belief that "natural melodic intervals" exist only insofar as successive spectrally complex sounds differing in fundamental frequency may have common spectral components [23] or a perceptual similarity due to pitch ambiguity [24, 14].

We hypothesize that the naturalness of certain melodic intervals has strongly influenced the construction of the musical scales that became culturally successful. In line with this hypothesis, the two scales for which final performance was best in experiment 1 (*Major_N* and *WT_S*) were the only scales endowed with cultural popularity: *WT_S* was initially unfamiliar to the participants, but it is essentially the Slendro scale of Javanese gamelan music (with a replication of the second degree one octave higher). Let us emphasize that perceptual sensitivity to frequency ratio simplicity in melodies may be strengthened by cultural factors even if it originally had a causal influence on these factors [25]. In other words, this sensitivity may be a product of *both* nature and culture.

From a certain point of view, our results are puzzling. Melodies based on scale *Major_N* can make use of intervals comprising any whole number of semitones (from 1 to 12). Thus, for *Major_N*, the detection of an error in melody 2 could not rest upon the detection of a specific interval. By contrast, for *WT_N*, an error could always be detected by noticing that melody 2 contained an interval comprising an odd number of semitones; most of

the time, an error actually produced two such intervals. Nevertheless, performance was poorer for *WT_N* than for *Major_N*, even after more than 3000 trials with each scale. This paradox shows that any model of perceptual processing assuming that the participants identified the intervals between successive notes of the melodies is unrealistic.

Instead, we propose a model based on the notes themselves (see Supplemental Figure S1). We estimated the statistical surprise elicited by notes from anomalous melodies, given a scale template, for the different experimental conditions. The reasoning was that higher surprise should lead to better detection of an anomaly. A Kullback-Leibler divergence was used to quantify surprise, taking into account the noise introduced during sensory encoding and memorization. As intended by the design of our experiments, the surprise elicited by individual notes could not explain the details of the behavioral results. However, an almost perfect fit to the data was obtained by sharpening the representation of scale templates around degrees corresponding to simple frequency ratios (Figure S1, right panel).

This leads to a resolution of the paradox pointed out above: during the task, presumably, participants did not compare the intervals of melody 1 and melody 2, nor did they compare the individual notes of melody 1 and melody 2. Instead, the model suggests that they perceived each note *in relation to the tonic* (lowest degree) of the implied scale, and monitored surprise relative to the internalized template of this scale. Then, because the internalized templates were especially accurate around simple frequency ratios, some melodies were more accurately processed than others.

The internalization of scale templates was likely in our experimental paradigm, with or without prior familiarity, as the scale template upon which a "normal" melody was built was used throughout multiple test sessions. Moreover, the first part of each trial (melody 1) revealed the tonic used for this trial. The effect of prior internal templates on current melody perception is also well documented [13, 26-29]. The existence of a special status for simple frequency ratios could be related to the phenomenon of harmonic fusion, i.e., the perceptual grouping of partials of spectrally complex periodic sounds [30-33]; internal representations of simple frequency ratios might be acquired already in utero as a result of exposition to vocal sounds [24]. Note, however, that simple frequency ratios are also expected to be perceptually special if pitch is physiologically computed in the temporal domain, by a mechanism akin to autocorrelation [34-36] or neural oscillations persisting after the stimulus [37].

In summary, the present work brings another perspective to the age-old mystery of the peculiar structure of musical scales: how melodies are naturally represented in human

memory may have contributed to the widespread use, over history and across cultures, of musical scales comprising both simple frequency ratios and a tonal center.

Methods

Musical scales employed in the test sessions

The eight scales employed (*Major_N*, *Major_S*, *WT_N*, *WT_S*, *Phrygian*, *Locrian*, *Phry* and *Loc*) can be described as sets of numbers representing the distance in semitones of each scale degree from the first (lowest) degree, a semitone being defined as a frequency ratio of $2^{1/12}$.

Using this convention,

Major_N = [0 2 4 5 7 9 11 12]: the major mode of the common diatonic scale;

Major_S = [0 2.4 4.8 6.0 8.4 10.8 13.2 14.4]: *Major_N* stretched by a factor of 1.2;

WT_N = [0 2 4 6 8 10 12]; a rarely used scale known as the "whole-tone" scale;

WT_S = [0 2.4 4.8 7.2 9.6 12 14.4]: *WT_N* stretched by a factor of 1.2;

Phrygian = [0 1 3 5 7 8 10 12]: the Phrygian mode of the common diatonic scale;

Locrian = [0 1 3 5 6 8 10 12]: the Locrian mode of the common diatonic scale;

Phry = [0 1 3 5 8 10]: a subset of *Phrygian*;

Loc = [0 1 3 6 8 10]: a subset of *Locrian*.

Sound examples for the different scales are available on ResearchBox:

https://researchbox.org/3561&PEER_REVIEW_passcode=NWGYMK

Stimuli and task

Our melodies consisted of equal-amplitude pure tones and were presented binaurally by headphones. Each tone had a total duration of 200 ms and was gated on and off with 20-ms cosinusoidal amplitude ramps. Consecutive tones in a melody were separated by 100-ms silent intervals. Two melodies ("melody 1" and "melody 2") were presented on each trial. They were separated by a 1-s silent interval. Melody 1 was obtained by randomly ordering the degrees of a given musical scale and selecting at random, between 370 and 740 Hz, the frequency of the tone corresponding to the first degree; the probability distribution underlying the frequency selection was rectangular, frequency being scaled logarithmically. When the scale upon which melody 1 was based was *Phry* or *Loc*, its first degree was represented twice, rather than only once, within melody 1 (which therefore consisted of 7 tones); the two occurrences of the first degree had independent and random temporal positions.

Melody 2 was obtained by reordering at random the component tones of melody 1, with or without a 1-semitone frequency shift in one tone. The probability of a shift was 0.5. In experiment 1, shifts in the lowest and highest component tones of melody 1 were prohibited but any other 1-semitone shift making the shifted tone out of scale (by more than 0.2 semitone) was possible. Thus there were 9 possible shifts when melody 1 was based on *Major_N* or *Major_S*, and 10 possible shifts for *WT_N* or *WT_S*. These 9 or 10 possible shifts were equiprobable. In experiments 2 and 3, there were only 4 possible shifts for each scale. These 4 equiprobable shifts are denoted by blue arrows in Figure 3A. They were either identical or symmetrical for the two scales compared in each experiment.

Following melody 2, the participant had unlimited time to indicate if this melody contained or not a sour note, i.e., a tone absent from melody 1. Responses were given by clicking on one of two virtual buttons displayed on the participant's computer screen, and were immediately followed by visual feedback: the button selected by the participant was colored in green if the response was correct, or in red if the response was wrong.

Organization of the test sessions

Each test session consisted of 220 trials with a fixed scale for melody 1: one block of 20 trials followed by five blocks of 40 trials. The initial block of 20 trials served as a warm-up and its results were ignored in the data analysis.

Experiments 1, 2, and 3 were respectively run on 30, 10, and 12 different days ("test days") for each participant (with one exception, see below). On each test day, two test sessions were run: one session for each of the two scales compared in the experiment (or experimental phase in the case of experiment 1). The order of the two scales was permuted from day to day. In addition, the order used on the first day was counterbalanced across participants. The two sessions of a given test day had to be separated by at least one hour. Another rule was that consecutive test days had to be separated by no more than two other days. In practice, the latter rule was occasionally transgressed by some participants; however, the interval never exceeded one week. For one participant in experiment 1, performance for *Major_N* was almost immediately quasi-perfect ($d' > 4$). Within Phase 1, consequently, this participant was tested in only four *Major_N* sessions, but 24 *WT_N* sessions. His data are ignored in Figure 2B (but not Table 1).

Due to the high number and demanding temporal schedule of the test sessions, it was difficult to run the experiments in laboratory conditions. This was done for only two participants in experiment 2. Except for these two cases, the participants tested themselves at

home, using their own computer – on which the compiled Matlab software running the experiment was initially installed – and generally their own headphones. High-quality headphones were lent to some participants without this equipment. At the end of each test day, the data collected on that day were transmitted to the experimenter.

Preliminary session

In each experiment, the test sessions were preceded by a preliminary session of about 45 min during which the participant and the experimenter were interacting orally. This session was primarily intended to familiarize the participant with the task and to provide detailed information on the frequency shifts to be detected. Familiarization was achieved by three blocks of 20 trials in which melody 1 was based on a simpler scale than those used in the test sessions. This scale was [0 2 5 7 9 12] in experiment 1, [0 3 5 8 12] in experiment 2, and [0 1 3 8 10] in experiment 3. A fourth block of 20 trials was run if, up to that point, performance had remained close to the chance level. Each participant was initially informed that, in this session as well as within each of the subsequent ones, melody 1 would be based on a fixed musical scale, the frequency shifts to be detected had a fixed size of 1 semitone, all of these shifts produced an out-of-scale tone, and this tone could occur at any temporal position within melody 2. Before the first block of trials, participants were also instructed to adjust the volume setting of their computer so as to hear the melodies at a moderate loudness level, and to use the same setting in each of the test sessions.

Data analysis

Performance was measured using the d' index of perceptual sensitivity from signal detection theory [38]: $d' = z(H) - z(F)$, z being the inverse of the normal cumulative distribution function, H the hit rate, and F the false-alarm rate. On some occasions, the obtained d' was infinite and thus had to be corrected. This was done by adding 0.5 to each of the four response counts (hits, misses, correct rejections and false alarms) [39]. Corrections occurred for 1.8 % (17 /956), 2.2 % (7/320), and 1.25 % (6/480) of the test sessions in experiments 1, 2, and 3, respectively. The full dataset is available on ResearchBox:

https://researchbox.org/3561&PEER_REVIEW_passcode=NWGYMK

Participants

The number of participants in each of the three experiments was *a priori* set to 16. In each experiment, some additional individuals were recruited but excluded after a few test sessions

because their performance remained close to the chance level ($d' < 0.5$) for both of the musical scales involved; there were two exclusions in experiments 1 and 3, and five exclusions in experiment 2. No one was tested in more than one experiment. The experiments were conducted following the guidelines of the INSERM Institutional Review Board (IRB00003888). Each participant provided written informed consent and was paid an hourly reward of 12 €/h.

We recruited young Western adults with self-reported normal hearing and a strong interest in music but at most a moderate level of musical education and expertise (professional musicians were deliberately avoided). Most of the participants were French students. Five of them (tested in experiment 1) had previously been tested in other experiments on auditory perception (but not melody perception); the remaining participants were all novices in this regard. Overall, the mean age was 22.4 y (23.4 y in experiment 1, 21.8 y in experiment 2, 22.0 y in experiment 3) and the age range 18-28 y; there were 13 males and 35 females (11 in experiments 1 and 2, 13 in experiment 3). In experiment 1, 8 participants identified themselves as amateur musicians; 10 participants had at least some ability to read a musical score while the remaining 6 had none. In experiments 2 and 3, respectively, 12 and 10 participants identified themselves as amateur musicians; only 8 and 9 were able to respond correctly to the question: what is a "minor third" in music?

Acknowledgements

This research was supported by a grant from the Agence Nationale de la Recherche (ANR-22-CE28-0023-03). We thank Alexandre Zénon for helpful discussions.

Author contributions

Conceptualization: LD; Methodology: LD, CS, DP; Software: LD, DP;
Formal analysis: LD, CS, DP; Investigation: LD; Writing – Original Draft: LD, DP;
Supervision: LD; Funding Acquisition: LD, CS, DP

Declaration of interests

The authors declare no competing interests.

Supplemental information

Figure S1

References

1. Savage, P.E., Brown, S., Sakai, E., and Currie, T.E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Nat. Acad. Sci. USA* 112, 8987-8992.
2. McBride, J.M., Passmore, S., and Tlusty, T. (2023). Convergent evolution in a large cross-cultural database of musical scales. *PLoS One* 18, e0284851.
3. Burns, E.M. (1999). Intervals, scales, and tuning. In Deutsch, D. (Ed.): *The Psychology of Music*, pp. 215-264 (San Diego, CA: Academic Press).
4. Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Percept.* 11, 289–305.
5. Sethares, W.A. (1998). *Timbre, Tuning, Spectrum, Scale* (New York: Springer-Verlag).
6. Gill, K.Z., and Purves, D. (2009). A biological rationale for musical scales. *PLoS One* 4, e8144.
7. Marjeh, R., Harrison, P.M.C., Lee, H., Deligiannaki, F., and Jacoby, N. (2024). Timbral effects on consonance disentangle psychoacoustic mechanisms and suggest perceptual origins for musical scales. *Nat Commun.* 15, 1482.
8. Demany, L., and Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *J. Acoust. Soc. Am.* 76, 57-66.
9. Cohen, A.J., Thorpe, L.A., and Trehub, S.E. (1987). Infants' perception of musical relations in short transposed tone sequences. *Canad. J. Psychol.* 41, 33-47.
10. Trainor, L.J., and Trehub, S.E. (1993). What mediates infants' and adults' superior processing of the major over the augmented triad? *Music Percept.* 11, 185-196.
11. Schellenberg, E.G., and Trehub, S.E., 1996. Natural musical intervals: evidence from infant listeners. *Psychol. Sci.* 7, 272–277.
12. Trehub, S.E. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669-673.
13. Shepard, R.N., and Jordan, D.S. (1984). Auditory illusions demonstrating that tones are assimilated to an internalized musical scale. *Science* 226, 1333-1334.
14. Parncutt, R. (2024). *Psychoacoustic Foundations of Major-Minor Tonality* (Cambridge, MA: MIT Press).

15. Trehub, S.E., Schellenberg, G., and Kamenetsky, S.B. (1999). Infants' and adults' perception of scale structure. *J. Exp. Psychol. Human Percept. Perform.* 25, 965-975.
16. Pelofi, C., and Farbood, M. (2021). Asymmetry in scales enhances learning of new musical structures. *Proc. Nat. Acad. Sci. USA* 118, e2014725118.
17. Loui, P., Wessel, D.L., and Hudson Kam, C.L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Percept.* 27, 377-388.
18. Loui, P. (2012). Learning and liking of melody and harmony: Further studies in artificial grammar learning. *Topics Cog. Sci.* 4, 1-14.
19. Leung, Y., and Dean, R.T. (2018). Learning unfamiliar pitch intervals: A novel paradigm for demonstrating the learning of statistical associations between musical pitches. *PLoS One* 13: e0203026.
20. Jacoby, N., Undurraga, E.A., McPherson, M.J., Valdés, J., Ossandón, T., and McDermott, J.H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Curr. Biol.* 29, 3229-3243.
21. Patel, A.D. (2008). *Music, Language, and the Brain*. Oxford University Press.
22. Lynch, M.P., Short, L.B., and Chua, R. (1995). Contributions of experience to the development of musical processing in infancy. *Dev. Psychobiol.* 28, 377-398.
23. Helmholtz, H. von (1877). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*, 4th ed. (Braunschweig, Germany: Vieweg).
24. Terhardt, E. (1974). Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* 55, 1061-1069.
25. Singh, M., and Mehr, S.A. (2023). Universality, domain-specificity and development of psychological responses to music. *Nat. Rev. Psychol.* 2, 333–346.
26. Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychol. Rev.* 85, 341–354.
27. Brattico, E., Tervaniemi, M., Näätänen, R., and Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Res.* 1117, 162-174.
28. Anglada-Tort, M., Harrison, P.M.C., Lee, H., and Jacoby, N. (2023). Large-scale iterated singing experiments reveal oral transmission mechanisms underlying music evolution. *Curr. Biol.* 33, 1472-1486.

29. Pressnitzer, D. (2023). Music cognition: Can you carry over that tune? *Curr. Biol.* 33, R296-R298.
30. Bendixen, A., Háden, G., Németh, R., Farkas, D., Török, M., and Winkler, I. (2015). Newborn infants detect cues of concurrent sound segregation. *Dev. Neurosci.* 37, 172-181.
31. Feng, L., and Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proc. Nat. Acad. Sci. USA* 114, E840-E848.
32. McPherson, M.J., Dolan, S.E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E.A., Jacoby, N., Godoy, R.A., and McDermott, J.H. (2020). Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nat. Commun.* 11, 2786.
33. Demany, L., Monteiro, G., Semal, C., Shamma, S., and Carlyon, R.P. (2021). The perception of octave pitch affinity and harmonic fusion have a common origin. *Hear. Res.* 404, 108213.
34. Licklider, J.C.R. (1951). A duplex theory of pitch perception. *Experientia* 7, 128-134.
35. Cariani, P.A., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones: I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
36. de Cheveigné, A. (2023). Why is the perceptual octave stretched? An account based on mismatched time constants within the auditory brainstem. *J. Acoust. Soc. Am.* 153, 2600-2610.
37. Large, E., Kim, J., Flaig, N., Bharucha, J., and Krumhansl, C. (2016). A neurodynamic account of musical tonality. *Music Percept.* 33, 319-331.
38. Green, D.M., and Swets, J.A. (1974). *Signal Detection Theory and Psychophysics* (Huntington, NY: Krieger).
39. Brown, G.S., and White, G. (2005). The optimal correction for estimating extreme discriminability. *Behav. Res. Meth.* 37, 436-449.

Figure legends

Figure 1: Illustration of the experimental task and predictions. **A.** The experimental task, illustrated for a trial in which melody 1 is based on the diatonic major scale. In the depicted case, the lowest and highest component tones of melody 1 correspond exactly to the note C. Participants had to indicate if melody 2 was "normal" or "anomalous". When melody 2 was "normal", its component tones were identical to those of melody 1 (but randomly reordered). The anomalies to be detected consisted of a one-semitone frequency shift creating an out-of-scale note. Here, this note (in red) is an F sharp, replacing the F of melody 1. **B.** Possible effects of practice on performance, when the musical scale on which melody 1 is based includes (scale X) or does not include (scale Y) simple frequency ratios. According to the "Familiarity" hypothesis, frequency ratio simplicity is beneficial in the task merely because it is immediately familiar, in consequence of musical enculturation. This hypothesis predicts an initial advantage of scale X, but also predicts that this advantage will be progressively abolished by practice, because practice makes scale Y more and more familiar. According to the "Naturalness" hypothesis, by contrast, frequency ratio simplicity is *intrinsically* favorable, due to universal properties of the human auditory system. An advantage of scale X over scale Y can then be found even after the saturation of perceptual learning due to practice.

Figure 2: Results of experiment 1. **A.** The four musical scales used in the test sessions: *Major_N*, *Major_S*, *WT_N*, and *WT_S*. Dots represent the scale degrees. Small-integer frequency ratios relative to the first degree are indicated by red vertical lines. **B.** Mean performance (d') across participants as a function of scale and session number. *Major_N* and *WT_N* were used in Phase 1 and Phase 3, while *Major_S* and *WT_S* were used in Phase 2. The error bars represent within-subject standard errors, computed with the Cousineau-Morey method. Each data point is the outcome of 3000 trials (200 trials of 15 participants).

Figure 3: Results of experiments 2 and 3. **A.** The musical scales used in the test sessions of experiment 2 (*Phrygian* and *Locrian*) and experiment 3 (*Phry* and *Loc*). As in Figure 2A, dots represent the scale degrees and red vertical lines position small-integer frequency ratios relative to the first degree. Blue arrows depict the frequency shifts that could occur from melody 1 to melody 2. For the two scales used in a given experiment, the shifts were either

identical or symmetrical. **B.** Results of experiment 2: mean performance (d') across participants as a function of scale and session number. The error bars represent (here and also in panels C and D) within-subject standard errors, computed with the Cousineau-Morey method. Each data point is the outcome of 3200 trials (200 trials of 16 participants). **C.** Same as B, but for experiment 3. **D.** Mean results of the four participants on whom experiment 3 was conducted twice. Each data point is the outcome of 800 trials.

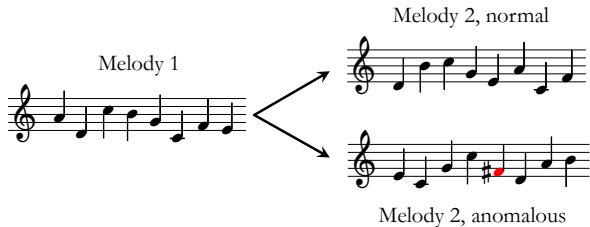
Table

Table 1: Final performance for the different scales. Final performance was taken as the mean d' value obtained in the last three test sessions. In the case of *Major_N* and *WT_N* (experiment 1), six d' values were averaged: those obtained in the last three test sessions of Phase 1 and Phase 3.

| Scale | Mean d' across participants | 95% confidence interval across participants |
|----------|-------------------------------|---|
| Major_N | 2.673 | ± 0.495 |
| Major_S | 0.446 | ± 0.113 |
| WT_N | 2.091 | ± 0.356 |
| WT_S | 2.519 | ± 0.474 |
| Phrygian | 2.658 | ± 0.505 |
| Locrian | 1.823 | ± 0.347 |
| Phry | 2.585 | ± 0.342 |
| Loc | 1.942 | ± 0.246 |

A

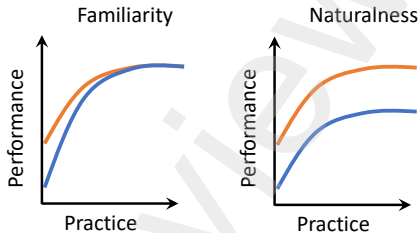
Experimental task

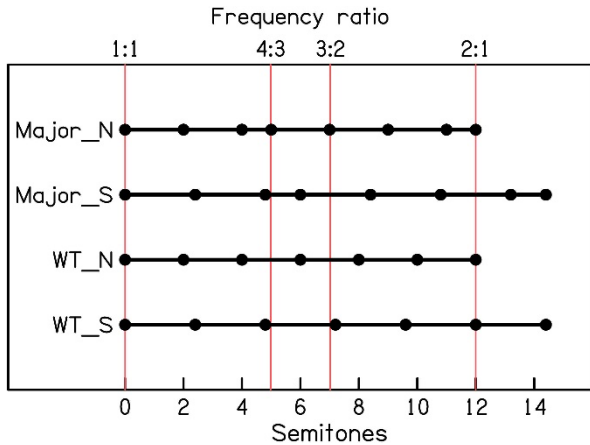
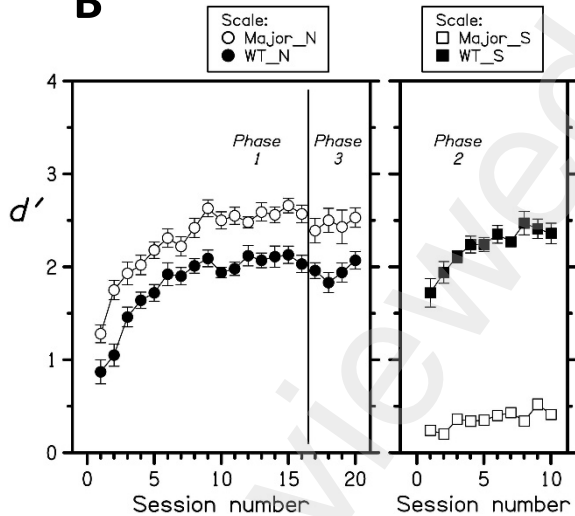
**B**

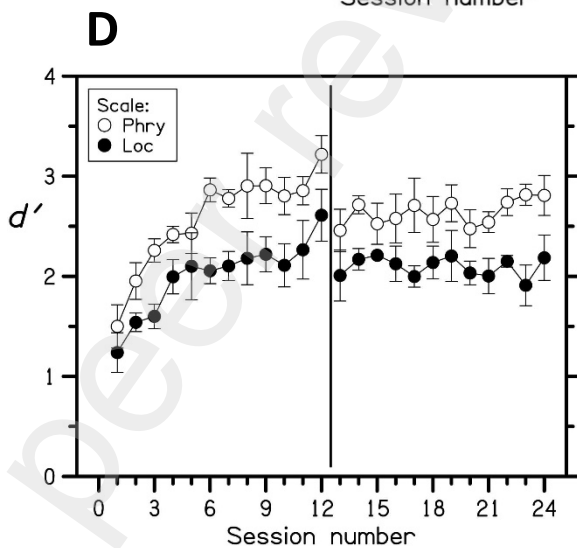
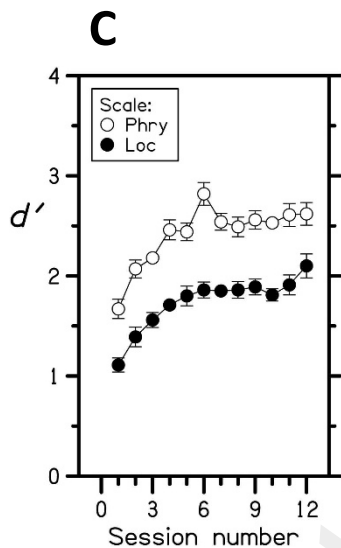
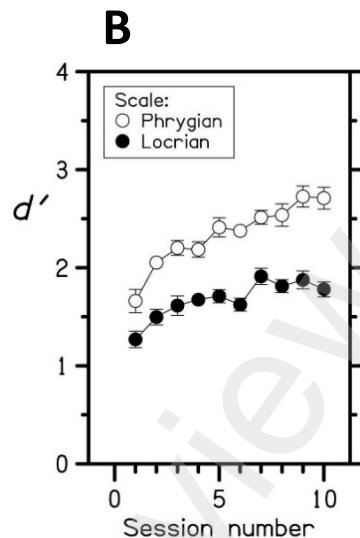
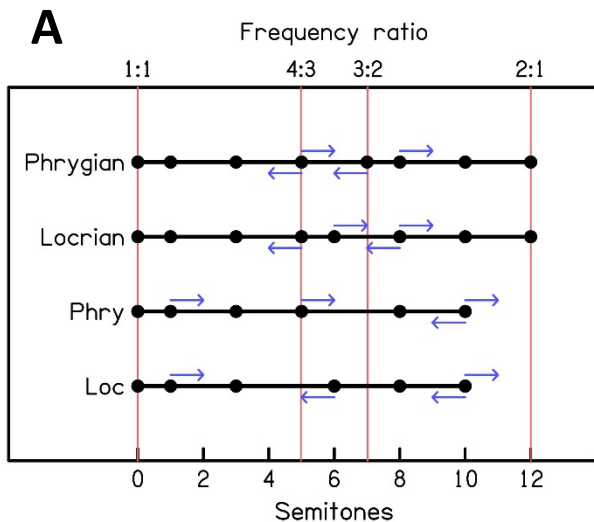
Predictions

Scale X (includes simple frequency ratios)

Scale Y (no simple frequency ratio)



A**B**



Supplemental Information

Simple frequency ratios naturally make precisely perceived melodies

Laurent Demany,¹ Catherine Semal,¹ and Daniel Pressnitzer²

¹Institut de Neurosciences Cognitives et Intégratives d'Aquitaine, CNRS, Université de Bordeaux, 33076 Bordeaux, France

²Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, CNRS, PSL University, 75005 Paris, France

Supplemental Figure S1

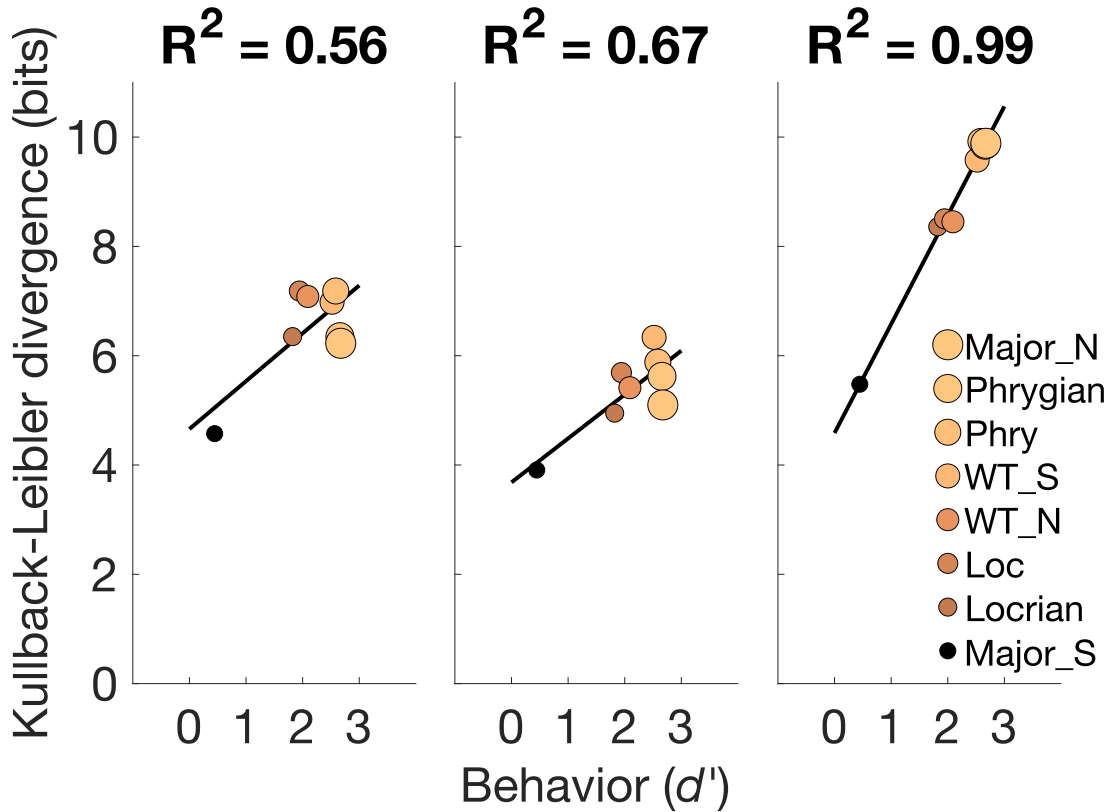


Fig. S1: Computational model. Estimated surprise (see below) and behavioral performance (data from Table 1). The eight musical scales used in the experiments are ordered by increasing levels of performance and identified by the color and size of the corresponding symbols.

To estimate surprise, we computed the Kullback-Leibler (KL) divergence, notated $D_{KL}(P_A||P_N)$, between the probability distribution of notes for melodies containing an anomaly, P_A , and the probability distribution of notes for the normal melodies, P_N . This divergence estimates the surprise resulting from using the template P_N to predict the observations P_A . To convert melodies into probability distributions, all possible notes and their relative frequency of occurrence during the experiment were identified and the resulting histogram was convolved with a Gaussian distribution of standard deviation SD_{noise} . The SD_{noise} parameter is intended to summarize the noise in sensory encoding and memorization of the pure tone frequencies constituting the melodies. A background, uniform non-zero probability was further added to avoid numerical errors (Matlab R2020a $eps = 2.22e-16$).

An optimization procedure (Nelder-Mead simplex, Matlab R2020a *fminsearch* with default parameters) maximized the correlation between the surprise estimates, $D_{KL}(P_A||P_N)$, and the behavioral results, d' . The optimization returned a value of $SD_{noise} = 0.09$ semitones, which produced a significant correlation between behavior and surprise ($R^2 = 0.56$; $p = 0.03$, left panel). However, this correlation was largely due to a very low predicted surprise for Major_S, which also produced very low behavioral performance. Omitting Major_S, the correlation disappeared ($R^2 = 0.05$, $p = 0.646$).

To quantify the hypothesis that a special status for simple intervals could account for the data, we modified the model in a single way: the representation of scale degrees was sharpened around simple intervals (1:1, 4:3, 3:2, and 2:1 frequency ratios) relative to the lowest degree. This was achieved by dividing the standard deviation of the Gaussian noise convolved with the scale degrees by a ratio, K_{sharp} , at simple intervals. Because of equal temperament and the use of stretched scales, scale degrees fell close but not exactly at simple intervals, so the sharpening ratio was computed from a continuous Gaussian function centered on simple intervals with standard deviation SD_{Ksharp} . Thus, notes at simple frequency ratios were represented with a standard deviation of SD_{noise}/K_{sharp} , while notes close to these ratios were represented with a standard deviation lying between SD_{noise}/K_{sharp} and SD_{noise} , depending on their distance from the simple ratios. To reduce the number of model parameters, we imposed $SD_{Ksharp} = SD_{noise}$. The optimization procedure was then performed on SD_{noise} and K_{sharp} . An improved correlation was found for the sharpened model ($SD_{noise} = 0.12$ st; $K_{sharp} = 1.4$; $R^2 = 0.67$; $p = 0.01$, middle panel). But again, this correlation was largely due to Major_S (excluding Major_S: $R^2 = 0.14$, $p = 0.41$).

Finally, we ran a simulation where only the P_N distributions were sharpened. This asymmetric model can be interpreted as simulating a more accurate representation of simple intervals for the normal scale template internalized in memory. Remember that melody 1, presented in the first part of every trial, was always normal and was always drawn from the same scale template throughout a test session. In the asymmetric model, the trial-to-trial observations of the scale of melody 2 were not sharpened. This new model produced an almost perfect fit to the data ($SD_{noise} = 0.12$ st; $K_{sharp} = 4.6$; $R^2 = 0.99$, $p = 2.7e-7$, right panel). The fit survived the omission of Major_S ($R^2 = 0.96$; $p = 1.5e-4$), showing that all the details of the experimental data were fully captured by the model.