AUTOMATIC COMPARISON OF GLOBAL CHILDREN'S AND ADULT SONGS SUPPORTS A SENSORIMOTOR HYPOTHESIS FOR THE ORIGIN OF MUSICAL SCALES

Shoichiro Sato[1], Joren Six[2], Peter Pfordresher[3], Shinya Fujii[1], Patrick E. Savage*[1]
[1]Keio University, Japan, [2]Ghent University, Belgium, [3]University at Buffalo, NY, USA
*Correspondence to: psavage@sfc.keio.ac.jp

ABSTRACT

Music throughout the world varies greatly, yet some musical features like scale structure display striking cross-cultural similarities. Are there musical laws or biological constraints that underlie this diversity? The "vocal mistuning" hypothesis proposes that cross-cultural regularities in musical scales arise from imprecision in vocal tuning, while the integer-ratio hypothesis proposes that they arise from perceptual principles based on psychoacoustic consonance. In order to test these hypotheses, we conducted automatic comparative analysis of 100 children's and adult songs from throughout the world. We found that children's songs tend to have narrower melodic range, fewer scale degrees, and less precise intonation than adult songs, consistent with motor limitations due to their earlier developmental stage. On the other hand, adult and children's songs share some common tuning intervals at small-integer ratios, particularly the perfect 5th (~3:2 ratio). These results suggest that some widespread aspects of musical scales may be caused by motor constraints, but also suggest that perceptual preferences for simple integer ratios might contribute to cross-cultural regularities in scale structure. We propose a "sensorimotor hypothesis" to unify these competing theories.

1. INTRODUCTION

Music exists in many different forms among almost all human societies, yet some common musical features are shared throughout the world. Although humans can discriminate more than 240 pitches within one octave, most tonal systems around the world incorporate scales with only seven or fewer pitches [1, 6, 12], but the precise tunings and combinations of these pitches used vary greatly throughout the world [5]. Theories of the origin of scales have been based on perceptual theories involving mathematical ratios since the time of Pythagoras [2]. Pitched notes on harmonic instruments produce multiple frequencies called harmonics or overtones, which resonate at integer multiples of the fundamental frequency (e.g., a note with a fundamental frequency of 100 Hz produces overtones at 200 Hz, 300 Hz, etc.). The intervals that have traditionally been considered "consonant" in Western music theory are also those with the simplest integer-ratios (e.g., 2:1 = octave, 3:2 = Perfect 5th, 4:3 = Perfect 4th; cf. Fig. 1).

Interval Name	Abbr.	Cents	Approx. frequency ratio	Example on keyboard
Perfect unison	**P1**	**0**	**1 : 1**	P1
Minor second	m2	100	16 : 15	m2
Major second	M2	200	9 : 8	M2
Minor third	m3	300	6 : 5	m3
Major third	M3	400	5 : 4	M3
Perfect fourth	**P4**	**500**	**4 : 3**	P4
Tritone	tt	600	7 : 5	tt
Perfect fifth	**P5**	**700**	**3 : 2**	P5
Minor sixth	m6	800	8 : 5	m6
Major sixth	M6	900	5 : 3	M6
Minor seventh	m7	1000	9 : 5	m7
Major seventh	M7	1100	15 : 8	M7
Perfect octave	**P8**	**1200**	**2 : 1**	P8

Figure 1. The equal tempered chromatic scale system. Boldface shows the diatonic ratios which have the simplest integer ratios.

However, these theories are generally based on tunable instruments that often use equal-tempered intervals that only approximate pure ratios (e.g., a pure perfect 5th is 702 cents, but a perfect 5th on a piano is produced at 700 cents), and some are skeptical as to whether this theory applies to vocal song, recognized as the most ancient and universal instrument of human music [11]. One alternative theory is that scales do not arise from perceptual constraints regarding integer ratios but instead due to production constraints on how precisely the voice can generate pitches [11]. This can be seen as a special case of the "motor constraint hypothesis", which proposes that many musical universals are not evolutionary adaptations for human music but simply byproducts of constraints on the way music is produced [13, 18]. The vocal mistuning theory argues that the universal tendency to use sparse scales with 7 or fewer scale degrees is because singing is too imprecise to allow

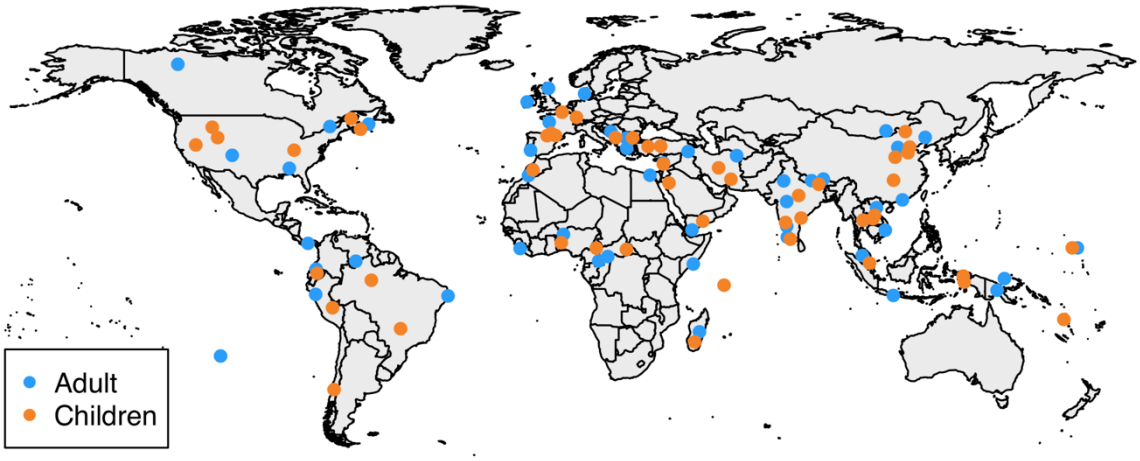

Figure 2. Map showing the approximate geographical distribution of 50 adult songs and 50 children songs used in this study

accurate production of scales using more than 7 scale degrees [11]. It predicts a negative correlation between tuning precision and number of scale degrees across musical genres. For example, children's songs should use sparser scales than adult songs because children are less able to precisely tune fine details of scales.

Although many studies of children's musical perception have been performed, there are only a handful of quantitative musicological studies of children's songs. In previous research, scales, melodic ranges, and other aspects of 100 children's songs from around the world were analyzed using the Cantometrics classification scheme [10, 15]. However, this study was limited to manual analysis and children's songs were not directly compared with adult songs.

In recent years, researchers have increasingly used large cross-cultural music corpora to address the relationship between human capacities and musical systems, as opposed to analyses that focus on specific musical cultures [14]. Thus, in the present study, we conduct automatic analyses to directly compare children's and adult's songs objectively using a matched global sample to examine the relationship between human vocal mistuning and musical scale structure.

In order to test the "vocal mistuning" hypothesis of scale origins, here we use an automatic method to compare children's songs and adult songs from around the world. Since children's vocal-motor control system are still developing, children should have smaller melodic ranges, less precise pitches, and sparser scales (i.e., fewer and more widely spaced scale degrees) than adult songs, even if both types of songs draw from the same overall tuning system (e.g., a given children's song might only utilize a sparser 4- or 5-note subset from within a broader 7-note tuning system used by adults). Meanwhile, in order to test the integer ratio hypothesis, we compared average scale degree tunings for children's and adult songs. If perceptual preferences for integer ratios contribute to scale structure, we predict common tunings on average across all songs, whereas if only motor constraints contribute to scale structure, we expect that no common trend will be found in this comparison.

2. Methods

2.1 The automatic pitch extraction tool - Tarsos

Conventional musical scale analysis has been dominated by manual transcription methods. Since different cultures have different tonal systems, it has generally been considered difficult to compare cross-cultural scale structures. Especially when transcribing non-Western folk song with Western annotation, it often happens that each pitch is fitted to a specific tuning system, such as equal-temperament, based on the perception the transcriber, which may not accurately reflect the original music. Therefore, in this study, we used the automatic pitch extraction tool Tarsos [16], because it was designed explicitly for automatic analysis of any scales from around the world without imposing any culture-specific theories. Tarsos has several pitch extraction modules such as pitch distribution filters, audio feedback tools, and scripting tools for batch processing of large databases of musical recordings for analyzing the pitch distribution.

2.2 Song Samples

In order to explore general tendencies of human music, we constructed a globally balanced sample of audio recordings (Fig. 2). First we chose 50 children's songs from *Le chant des enfants du monde* [4], the "Lullabies and Children's Songs" CD from the *UNESCO Collection* [17], and *Mama Lisa's World International Music & Culture* [19], selecting 5-6 songs each from the nine regions designated by the *Garland encyclopedia of world music* [9] (Africa, South America, North America, Southeast Asia, South Asia, East Asia, Middle East, Europe, and Oceania). After selecting children's songs, we used the same regional sampling process to select 50 traditional adult songs from the *Garland Encyclopedia of World Music* [9], *UNESCO Collection of Traditional Music* [17] and *Folkways*

Records that were as geographically matched as possible with the children's songs.

Most samples from these CD sets were recordings by ethnomusicologists of folk songs transmitted by oral tradition in relatively small societies that were minimally influenced by Western music. "Adult songs", as defined here, are songs sung by adults, while "children's songs" are songs sung by children (generally 8~16 years old), including nursery rhythms and gaming songs, but not lullabies or other genres sung by adults to children. Gender differences, musical functions, performance context, etc. were not considered for this study.

Before choosing samples from the CD sets, we manually checked Tarsos's automatic pitch extraction by ear in order to determine whether the songs work sufficiently with pitch detection; usable songs were added to the corpus and songs with instrumental accompaniment, polyphonic singing, or heavy background noise were excluded due to the inaccuracy of the pitch extraction algorithm on polyphonic recordings.

2.3 Comparison of melodic range measurements

First, we measured the melodic ranges to check how much vocal development affects music expression. Melodic range was calculated based on Tarsos's pitch annotation output function by subtracting the lowest pitch from the highest pitch to appear in the song, manually excluding transcription errors due to noise or overtones from the calculation (Fig. 3).

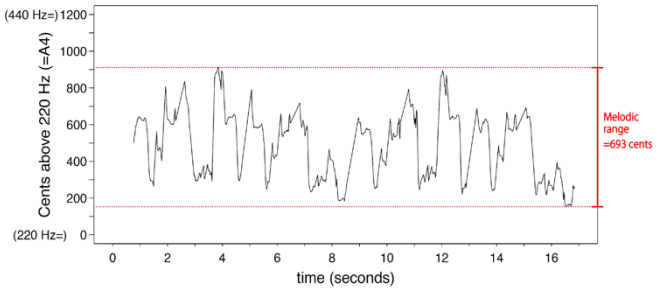

Figure 3. An example of melodic range analysis for an excerpt from "Plou i fa so" from Catalogne, Spain.

2.4 Comparison of number of scale degrees and vocal imprecision

2.4.1 *Number of Scale Degrees*

We examined the number of scale degrees to investigate how pitch precision influences scale structure. Tarsos first extracts the pitch histogram by detecting each pitch and number of annotations from the recording in real time. Next, another pitch distribution filter combines this pitch histogram across octaves to create a pitch class histogram, visualizing how pitch classes are distributed within one octave. This is expressed in cents [5] ranging from 0 to 1200 (see Fig. 1). Peak picking (see Fig. 4) is performed almost automatically, yet in order to enable extraction with the highest accuracy for all songs, we manually adjusted the parameter values for window (minimum distance between two peaks) and threshold (minimum occurrence necessary to be considered a "peak") depending on the songs. In order to automate these processes, we plan to evaluate optimal window and threshold values to allow objective peak picking without the need for manual adjustment in the future.

We used Tarsos's default YIN pitch estimation algorithm in this analysis [16]. Figure 4 shows examples of a children's song that has a relatively imprecise distribution, and an adult song that has more precise peaks.

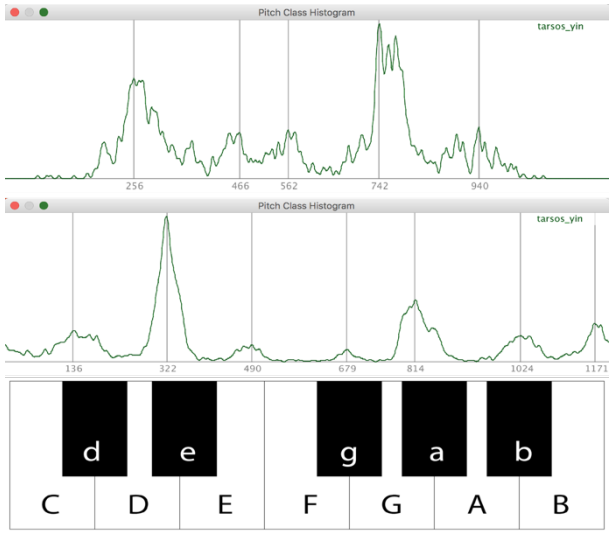

Figure 4. Automatic analysis of traditional West African scales: "Sigereti Fe Bara" (top, children's song, Pentatonic) and "Vai Call to Prayer" (bottom, adult song, Heptatonic). Vertical lines represent automatically detected scale degrees. The horizontal axis of the pitch class histogram shows pitch class across one octave (0 to 1200 cents) and the vertical axis shows the relative frequency of annotations.

2.4.2 *Calculating vocal imprecision*

We developed a novel formula (1) for quantifying the vocal imprecision (*I*). After obtaining pitch class histogram data, we took a width of ± 50 cents from each peak (*P*) and got the intersecting points (qi_a, q_{ib} represent the frequency of occurrence for pitch classes that span a quarter note (50 cents) above and below each peak (P_i), respectively. First, we calculated the average imprecision of each peak by dividing the frequency a quarter tone away from the peak by the frequency of occurrence for the peak pitch class; $((qi_a + q_{ib}) / 2 / P)$ and then averaged this imprecision across all pitch classes, such that 1 represents maximum imprecision (essentially no peaks), while 0 represents maximum precision (where scale degrees never fall more than one quarter tone from the peak; see Fig. 5 for details). We performed this calculation for all 100 songs and evaluated the correlation between tuning imprecision and number of scale degrees (Fig. 8C).

$$I = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{q_{ia}+q_{ib}}{2}\right)\Big/ P_i \qquad (1)$$

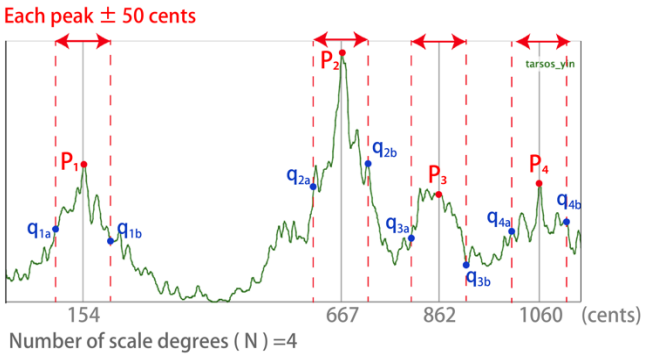

Figure 5. Visualization of formula for calculating imprecision, using an example children's song (see text for details).

2.5 Comparison of normalized scale analysis

Finally, the scale normalization to the tonal center (= tuning into same key) was performed to verify whether there is a tendency toward common intervals relative to the tonal centre of the scales. In order to compare scales across cultures, we attempted to normalize each scale by tuning the final pitch class to be 0 cents, following the method attempted in our previous study [7]. When the ending note is not detectable because of fade out or excerpts in a song, we instead normalized to the most frequent note as a tonal center (recent analyses [3] suggest that there is little difference between these two methods, leading us to propose consistently normalizing to the most frequent note in the future). Figure 6 shows an example of two normalized pitch class histograms. As shown in the figure, the tonal center was normalized to 0/1200 cents. All counts of pitch annotations are calculated as a percentage of pitch frequency, so as to be able to compare across recordings regardless of recording length.

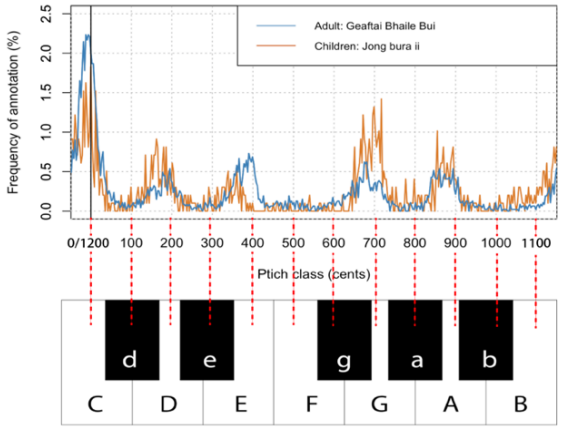

Figure 6. Examples of pitch class histograms for a children's song (Indonesian folk song "*Jong bura ii*", orange) and an adult song (Irish folk song "*Geaftai Bhaile Bui*", blue). Pitch class histograms are normalized so that the final/most frequent note is set to 0 cents. Both normalized pitch class histogram demonstrates similar scale structure:

3. RESULTS

3.1 Comparison of melodic range measurement

Figure 7 shows the distribution of melodic range measurements for 100 children's and adult songs. The mean absolute melodic range of the 50 children's songs was 941 cents (i.e., more than a minor 6th), while that of the 50 adult's songs was 1362 cents (i.e., more than a minor 9th): almost 50 % greater than children's songs ($t(98) = 5.8$, $p = 1.2\text{x}10^{-7}$). This result suggests that developmental constraints on vocal production influence musical expression. While children's songs tended to fall within one octave (~ 1200 cents), some adult songs showed a range of up to two octaves (~ 2400 cents). The average of a minor 9th range for adult songs is surprising given that previous manual research using a similar sample found that adult music throughout the world consistently tended to have a range of less than one octave [12]. This may be due to our automated analysis method overestimating melodic range due to being overly sensitive to outliers or to octave errors in transcription. In the future we intend to explore other algorithms besides the YIN algorithm and other measures that are less sensitive to outliers, such as interquartile range.

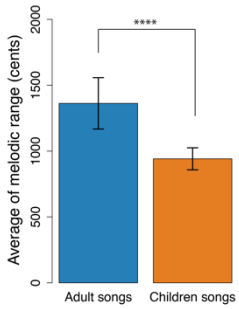

Figure 7. Comparison of average melodic range for children's and adult songs. Error bars = 95% confidence intervals.

3.2 Number of scale degrees and vocal imprecision

Figure 8A shows the average of the number of scale degrees for children's and adult songs. The scales of all songs were composed of seven or fewer scale degrees, except for one adult song with an 8-note scale. The use of pentatonic scales was the most dominant, accounting for approximately 1/3 of all songs (36% of adult corpus, 30% of children corpus). The average number of scale degrees in adult sample of 5.8 was significantly higher than that of the children's sample of 4.5 ($t(98) = 1.98$, $p = .007$; Fig. 8A). The mean vocal imprecision for children's songs of 0.41 was significantly higher than the average imprecision for adult songs of 0.34 ($t(98)=1.96$, $p=.009$; Fig.8B). There was a negative correlation between the number of scale degrees and vocal imprecision ($r=-0.23$, $p=.001$; Fig. 8C).

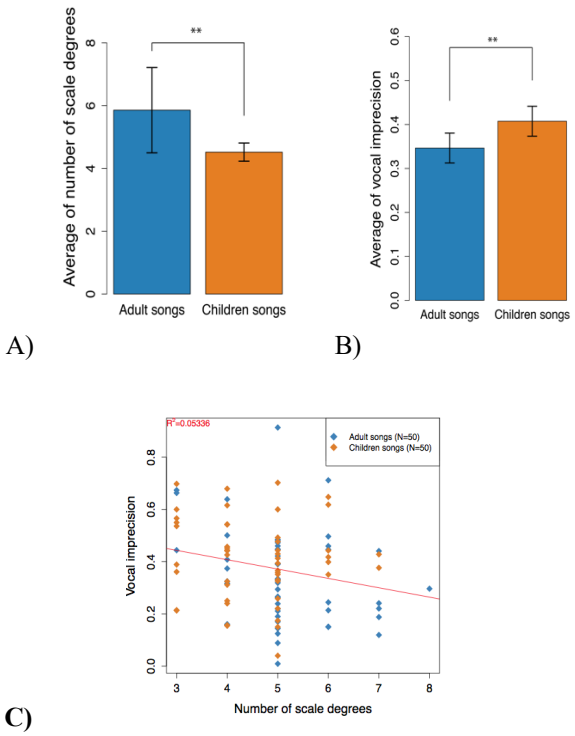

Figure 8. (A) The comparison of average number of scale degrees between children and adult songs. (B) The comparison of average vocal imprecision between children and adult songs. (C) Scatter plot and regression line showing the relationship between the number of scale degrees and vocal imprecision between children songs and adult songs. Error bars = 95% confidence intervals.

3.3 Normalized scale analysis

Figure 9 shows average scale tunings across children's and adult songs. Both children's and adult songs displayed peaks at the same four approximate intervals: perfect 5th (3:2 ratio, ~700 cents), perfect 4th (4:3 ratio, ~500 cents), major 3rd (5:4 ratio, ~400 cents), and major 2nd (9:8 ratio, ~200 cents), although the peaks were more precise for the adult songs. This result suggests that some aspects governing scale structure are consistent regardless of developmental stage. Note that tonality was not used as a criterion in sampling songs, so the predominance of major thirds over minor thirds reflects that major tonalities are more common cross-culturally.

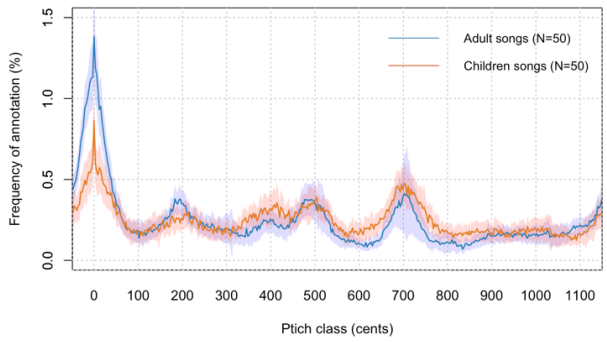

Figure 9. Average scale tunings across children's and adult songs. The x-axis begins and ends at 1150 cents in order to show the distribution around the tonal center (set to 0 cents). The transparent portions along the line of the pitch class histogram represent the 95% confidence intervals.

4. DISCUSSION AND FUTURE WORK

4.1 Discussion

These comparative analyses measuring melodic range, number of scale degrees, imprecision, and scale tuning, were conducted to test the hypothesis that cross-cultural regularities in pitch structure are determined by developmental constraints, particularly regarding vocal mistuning. We found that adult songs throughout the world consistently employ wider melodic ranges and denser scales with more precise tuning than children's songs, as predicted by the vocal mistuning hypothesis [11]. However, we also found similar small-integer ratio intervals appear consistently in both children's and adult songs, which is predicted by perceptual hypotheses based on small-integer ratio consonance [2, 6] but not by motor constraint hypotheses such as the vocal mistuning hypothesis.

Based on these results, we believe neither the perceptually-based integer ratio hypothesis nor the production-based vocal mistuning hypothesis alone fully explains cross-cultural regularities in scale structure [11]. We instead propose a more nuanced "sensorimotor hypothesis" for the origin of musical scales that combines these previous two hypotheses. This sensorimotor hypothesis argues that scale structure is determined by a balance between optimizing interval size for accurate production and optimizing ratios among scale degrees for maximal consonance in group music-making. Among other things, this hypothesis predicts that motor constraints will play a stronger role in governing solo and monophonic music, while perceptual constraints will play a larger role in polyphonic and group music.

4.2 Future work

In future studies, we plan to expand the scope of our methods to better test our sensorimotor hypothesis and other hypotheses for the origins of musical structure. This includes expanding the sample of human songs as well as including instrumental music, speech, and animal song for comparison [8]. To do so, we need to refine and further automate our process (e.g., peak-picking, noise removal for melodic range analysis, quantification of imprecision, improved automatic pitch extraction to accommodate polyphonic music) to be able to analyze larger samples while making fewer manual judgments. It will also be necessary to test our predictions through cross-cultural and cross-species behavioral experiments at different developmental stages in addition to corpus studies such as this.

Comprehensively addressing such improvements will require substantial investment to go beyond the current state-of-the-art in musical information retrieval, music cognition, and ethnomusicology, but hold the promise for

understanding how and why music has evolved to hold such universal power, and how we might harness that power for a more harmonious future.

5. AUTHOR CONTRIBUTIONS

S.S., P.E.S., S.F., J.S., and P.Q.P. designed the study; S.S. analyzed the data, supervised by P.E.S.; S.S. and P.E.S. drafted the manuscript.

6. ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Young Scientists from the Japan Society for the Promotion of Science, Keio Research Institute at SFC Startup Grant, and a Keio Gijuku Academic Development Fund grant to P.E.S.

7. REFERENCES

[1] Brown, A., Jordania, J. (2013). Universals in the world's musics, *Psychology of Music*, *41*(2), 229–248.

[2] Bowling, L. D., Purves, D., & Gill, Z. K. (2018). Vocal similarity predicts the relative attraction of musical chords. *Proceedings of the National Academy of Sciences of the U. S. A.,115*(1), 216-221.

[3] Chiba, G., Ho, M.-J., Sato, S., Kuroyanagi, J., Six, J., Pfordresher, P. Q., Tierney, A. T., Fujii, S., & Savage, P. E. (2019). Small-integer ratio scales predominate throughout the world's music. *PsyArXiv* preprint. https://doi.org/10.31234/osf.io/5bghm

[4] Corpataux, F. (1993-2018). *Le Chant des enfants du monde* [57 CDs]. ARION.

[5] Ellis, J. A. (1885). On the musical scales of various nations. *Journal of the Society of Arts, 23*(1688), 435-527.

[6] Gill, Z. K., & Purves, D. (2009). A biological rationale for musical scales. *PLOS ONE*, 4(12), e8144.

[7] Ho, M.-J., Sato, S., Kuroyanagi, J., Six, J., Brown, S., Fujii, S., & Savage, P. E. (2018). Automatic analysis of global music recordings suggests scale tuning universals. *Extended Abstracts for the Late-Breaking Demo Session of the 19th International Society for Music Information Retrieval Conference*.

[8] Kuroyanagi, J., Sato, S., Ho, M.-J., Chiba, G., Six, J., Pfordresher, P. Q., Tierney, A. T., Fujii, S., & Savage, P. E. (2019). Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales. *Proceedings of the 9th International Workshop on Folk Music Analysis*. Preprint: https://doi.org/10.31234/osf.io/zpv5w

[9] Nettl, B., Stone, R., Porter, J., & Rice, T., eds. (1998–2002). *The Garland encyclopedia of world music* (Garland, New York).

[10] Pai, J. S. (2009). *Discovering musical characteristics of children's songs from various parts of the world.* MA thesis. University of British Columbia.

[11] Pfordresher, Q. P., & Brown, S. (2017). Vocal mistuning reveals the origin of musical scales. *Journal of Cognitive Psychology*, *29*(1), 35–52.

[12] Savage, P. E., Brown, S, Sakai, E., & Currie, E. T. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy.of Sciences of the U. S. A.*, *112*(29), 8987–8992.

[13] Savage, P. E., Tierney, T. A., & A. D. Patel, (2017). Global music recordings support the motor constraint hypothesis for human and avian song contour. *Music Perception*, *34*(3), 327–334.

[14] Savage, P. E., & Brown, S., (2013). Toward a new comparative musicology. *Analytical Approaches to World Music*, *2*(2), 148–197.

[15] Savage, P. E. (2018). Alan Lomax's Cantometrics Project: A comprehensive review," *Music & Science.*, 1, 1–19, https://doi.org/10.1177/2059204318786084.

[16] Six, J., Cornelis, O., & Leman, M. (2013). Tarsos, a modular platform for precise pitch analysis of Western and non-Western music. *Journal of New Music Research*, *42*(2), 113-129.

[17] Smithsonian Institution Archives Record Unit (1961-2006), "*UNESCO Collection of traditional music*" [123 CD set].

[18] Tierney, T. A., Russo, A. F., & Patel, D. A. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences of the U. S. A.*, *108*(37), 15510–15515.

[19] Yannucci, L. (2019). "*Mama Lisa's World International Music & Culture*". https://www.mamalisa.com.