

Running head: HUMAN AND AVIAN SONG CONTOUR

Global music recordings support the motor constraint hypothesis for human and avian
song contour

Patrick E. Savage

Tokyo University of the Arts, Japan

Adam T. Tierney

Birkbeck University of London, UK

Aniruddh D. Patel

Tufts University, USA

Abstract

There has recently been renewed interest in using quantitative data to explore questions about musical universals. One explanation for certain musical universals is that they reflect ways of singing that are most energetically efficient, as opposed to biological specializations for human music. Previous research found support for this "motor constraint hypothesis" by comparing pitch contour shapes in samples of human and avian songs, but the sample of human songs was limited to notated scores of European and Chinese folk songs from the Essen database. Here we attempt to test this hypothesis using a more diverse global sample of human music recordings from the *Garland Encyclopedia of World Music*. By directly comparing pitch contour shapes in a diverse sample of human songs and bird songs, we found that both human and bird songs tend to employ similar descending/arched melodic contours despite substantial differences in absolute pitch and duration. This preference was consistent for both Western and non-Western songs. Surprisingly, we also found that the global samples of human and bird song contours were significantly more correlated with one another than either was with the Essen contours. Our findings of broad cross-cultural and cross-species parallels support the motor constraint hypothesis for melodic contour. More generally, our findings demonstrate the importance of greater collaboration between ethnomusicology and music psychology.

Keywords: Cross-cultural, melodic contour, audio analysis, music evolution, bird song

1. Introduction

The 21st century has seen a resurgence of interest in cross-cultural musical universals as part of a broader renewal of interest in music and evolution (Brown & Jordania, 2013; Honing, Cate, Peretz, & Trehub, 2015; Wallin, Merker, & Brown, 2000). The existence of such universals has been disputed for decades, but a recent study provided quantitative evidence for dozens of statistical universals in the world's music (Savage, Brown, Sakai, & Currie, 2015). These universals spanned a variety of domains including pitch (e.g., 7 or fewer scale degrees, descending/arched melodic contours), rhythm (e.g., meters based on multiples of 2 beats), form (e.g., short phrases lasting less than 9 seconds), instrumentation (e.g., percussion associated with group performance), performance style (e.g., chest voice), and social context (e.g., predominance of performances in groups and by males). Savage et al. speculated on a variety of explanations for these universals, but admitted that future research involving cross-species and other direct comparisons would be needed to test these explanations.

Cross-cultural universals have been of interest to evolutionary psychologists as candidates for biological adaptations (Barkow, Cosmides & Tooby, 1992). On the other hand, some universals may have non-adaptive explanations. For example, both language and music are cross-culturally universal, but many authors argue that evolution has biologically specialized our brains for language processing, while music arose as an invention based on other cognitive abilities, including language (Patel, 2008; 2010; Pinker, 1997).

One non-adaptive explanation for certain musical universals is Tierney, Russo and Patel's (2011) "motor constraint hypothesis". This hypothesis proposes that certain

widespread features of music reflect energetically efficient use of the vocal motor production system rather than evolutionary adaptations specific to human music. For example, Tierney et al. (2011) proposed that the predominance of arched and descending melodic contours in vocal music (Sachs, 1962) may reflect the tendency for air pressure beneath the vocal folds ("subglottal pressure") to increase rapidly at the start of a continuous vocalization and then decline gradually over the course of the vocalization (cf. the concept of "declination" in linguistics; Ladd, 1984; Slifka 2006). The biomechanics and physics of vocalization do not prevent humans from ending melodic contours on higher pitches, since humans can control the tension of their vocal folds independently of subglottal pressure. (In other words, biomechanics or physics alone does not predict a bias toward certain melodic contour shapes in human song.) However, all else being equal, higher subglottal pressure facilitates faster rates of vocal fold vibration and thus higher pitches, which should make it energetically more efficient to produce arched and descending pitch contours than contours with opposite shapes (i.e., U-shaped and ascending contours). To test this hypothesis, Tierney et al. compared pitch contour shapes in large corpora of human and bird songs, since birds and humans share many features of peripheral sound production (Elemans et al., 2015). As predicted by the motor constraint hypothesis, both human and bird song contours showed a bias toward descending and arched contours rather than ascending and U-shaped contours.

While the studies of Tierney et al. (2011) and Savage et al. (2015) both addressed patterns in pitch contour shapes in melodies, certain limitations make it difficult to compare their findings. Tierney et al.'s sample of bird songs was based on audio recordings from around the world, but their musical corpus was restricted to notated

scores of European and Chinese folk songs from the Essen folksong database (Selfridge-Field, 1995). Savage et al.'s musical sample had a much more global scope, but pitch contour analysis was done by ear using only three possible classification options for the most common contour of an entire *song* ("arched/descending", "U-shaped/ascending", or "other"). This stands in contrast to Tierney et al.'s approach, which analyzed each individual *breath group* (e.g., phrases for human songs) using automated contour analysis. These limitations raise the question of how the contours of human and bird songs might compare if the same type of automated contour analysis were applied to a worldwide sample of actual recordings of human songs.

Our goal for the present study was to extend and synthesize the findings of Tierney et al. (2011) and Savage et al. (2015) regarding human and avian song contour. By applying the fine-grained automated analytic technique used for bird song by Tierney et al. to the global music corpus used by Savage et al., we hoped to test the motor constraint hypothesis for song contour using a uniquely detailed cross-cultural and cross-species comparison.

2. Method

2.1. Materials

Our novel dataset consisted of 387 phrase contours from 35 human songs (Fig. 1). (Metadata for these 35 songs can be found in Savage et al.'s [2015] SI Dataset, recordings no. 4, 15, 25, 37, 40, 67, 72, 74, 80, 86, 89, 116, 143, 151, 152, 155, 164, 167, 168, 175, 190, 206, 214, 218, 219, 221, 224, 226, 228, 229, 247, 257, 263, 269, and 294). These were obtained by taking all solo a cappella sung phrases from the 304 audio recordings from the *Garland Encyclopedia of World Music* (Nettl, Stone, Porter, & Rice, 1998),

excluding phrases whose pitches could not be automatically tracked due to high levels of background noise. Sixteen of these songs (hereafter designated the "Western" subsample; n=208 phrases) were from Europe or were performed by speakers of European languages, while the remaining 19 were not (the "non-Western" subsample; n=179 phrases). Phrases were operationally defined as periods of continuous vocalizing separated by breaths, as annotated by ear by an assistant who was blind to our hypotheses.

We recognize that 387 phrases and 35 songs is a relatively small sample of human music, but this sample has the rare and important property of having been chosen by ethnomusicologists to represent the diversity of human music. Thus we strongly suspect that the pitch contour biases we observe in this dataset will replicate when larger culturally-diverse corpora of singing are studied in the future.

<INSERT FIG. 1 APPROXIMATELY HERE>

Our human song data was compared against birdsong and human song data published previously by Tierney et al. (2011). This birdsong data consisted of the pitch contours of 1,094 notes from 80 recordings of taxonomically-diverse songbirds (recordings were obtained from a variety of sources including the Cornell Laboratory of Ornithology, the British Library, and the Borror Laboratory of Bioacoustics). For birds, who are capable of taking 'minibreaths' between each note (Suthers, Goller & Pyte, 1999), contours were extracted from continuous vocalizations separated by either silence (97% of sample) or sudden jumps in pitch (3%; see Tierney et al. 2011, Fig. 2 for examples). Tierney et al.'s human song data consisted of 52,899 phrase contours from 9,467 folk song notations from the Essen database (Selfridge-Field, 1995) - a much larger

but also much less cross-culturally balanced sample than the *Garland Encyclopedia of World Music*.

2.2. Procedure

We followed the same analysis method for human song contour as Tierney et al. (2011) performed for bird song contour (Fig. 2). Briefly, the fundamental frequency throughout each breath group (i.e., phrase for human song, note for bird song) was extracted from spectrograms. Human song pitch data were tracked via autocorrelation in Praat¹. Bird song pitch data were tracked using an automatic spectral contour detection algorithm in SIGNAL version 3.1 (Engineering Design), as the pure-tonal nature of the majority of the birdsongs caused spectral contour detection to be more accurate than autocorrelation. For both human and bird songs, pitch contours were carefully checked and manually corrected where necessary.

Each frequency contour was then converted from a linear Hz scale into pitch contours in a normalized logarithmic semitone scale using the formula $ST = 12 \cdot \log_2(F/\text{mean}(F))$, where F is the frequency of a data point in Hz and $\text{mean}(F)$ is the mean frequency of the contour. Note from the above equation that this does not convert continuous frequencies into discrete pitch classes, it simply transforms them into a continuous semitone scale that still preserves any microtonality that may be present. Average pitch–time contours were generated by sampling each normalized pitch contour at 50 equally spaced time points, then averaging across contours.

¹ Pitch floor: 75 Hz, number of candidates: 15, silence threshold: 0.03, voicing threshold: 0.8, octave cost: 0.01, octave-jump cost: 0.35, voiced/unvoiced cost: 0.14, pitch ceiling: 600 Hz.

<INSERT FIG. 2 APPROXIMATELY HERE>

A difference between our analysis and Tierney et al.'s was that we chose to focus on the average contours (as described above), whereas Tierney et al. primary analysis classified each individual contour into one of nine possible shapes: "arched", "U-shaped", "descending", "ascending", "descending-horizontal", "ascending-horizontal", "horizontal-descending", "horizontal-ascending", and "horizontal" (following Huron 1996, though note that in a subsidiary analysis Tierney et al. computed average contours as described above: see the supporting information in Tierney et al., 2011). Classification of contours into one of nine shapes seemed insufficient for the types of complex contours seen in our global sample (e.g., the third contour in Fig. 2 might be best-described as "M-shaped" or "undulating", but is [arguably] misclassified by Huron's 9-shape scheme as "U-shaped").

Thus, we decided to focus our analysis instead on the more-nuanced average contours (see Fig. 3). To test whether these average contours were significantly arched/descending as predicted by the motor constraint hypothesis, we fit a quadratic model ($y = ax^2 + bx + c$) to the median values for each of the 50 time-points, and then performed a Monte Carlo analysis in which the 50 time points for each song phrase were randomized with 1,000 permutations to test whether the observed values of a (curve) and b (slope) were significantly lower than this random distribution.

3. Results

Fig. 3 shows the averaged overall contours for the Garland song corpus, including Western and non-Western sub-samples. Tierney et al.'s previously published data from

the bird song corpus and the Essen folksong corpus are also shown for comparison (reproduced from the supporting information of Tierney et al., 2011).

<INSERT FIG. 3 APPROXIMATELY HERE>

As predicted, the average contour of human songs in the Garland corpus displayed contours that were significantly arched ($a = -0.0017, p < 0.001$) and descending ($b = -0.046, p < 0.001$), indicating the presence of both a gradual drop in pitch as well as an initial rise and a final steep fall. These patterns were consistent in both the Western and non-Western subsamples (all $p < .001$)², and matched the ones found previously by Tierney et al. (2011) for bird songs and for both the European and the Chinese subsets from the Essen database.

Although the key predictions regarding descending/arched vs. ascending/U-shaped contours were consistent for all human and bird song samples and sub-samples, there were major differences in absolute pitch and duration values, and minor differences in the overall contour patterns. Compared to human song phrases, bird song notes on average had much higher pitch, smaller pitch ranges, and much shorter durations (Table 1). Bird songs also displayed a more gradual initial rise compared to the sudden rises characterizing all human contour profiles.

<INSERT TABLE 1 APPROXIMATELY HERE>

Surprisingly, the average contours in our global sample of human songs are actually significantly more correlated with those in Tierney et al.'s sample of bird songs than with those in their sample of European/Chinese songs from the Essen database ($r =$

² Western: $a = -0.0018, b = -0.0645$; Non-western: $a = -0.0017, b = -0.030$.

.95 vs. $r = .88$; $p < .01$; Fig. 3)³. This unexpected finding appears to be due to the more symmetrical nature of the Essen database contours, which more closely approximate an arch by staying higher for more of the phrase. In contrast, although the bird song contours don't show as dramatic an initial rise as the human song contours, the global samples of human and bird song contours both peak earlier and fall more steadily than the Essen contours.

4. Discussion

By using automated audio contour analysis of a global sample of human songs, we were able to confirm some of the key predictions of the motor constraint hypothesis for human and avian song contour, as well as to find novel cross-cultural/cross-species similarities and differences. Despite major differences in absolute pitch, melodic range, and duration, our analysis showed substantial similarities in the pitch contours of human and bird songs. As predicted, the average contour shapes tended toward descending/arched contours rather than ascending/arched contours.

The motor constraint hypothesis predicts a statistical bias toward melodic contours that are descending/arched because such contours are more energetically efficient to produce, given the increases and decreases in subglottal pressure over the course of a vocalization. While our results are consistent with this hypothesis, they do not rule out alternative explanations. For example, complex vocal learning has evolved convergently in humans, songbirds, and a few other groups of animals. Vocal learning may lead to certain other similarities in musical capacities or behaviors (Honing et al.,

³ Paired correlation test using *psych* package in R V3.2.3; bird song-Essen correlation: $r =$

2015; Patel, 2014). Thus, it is conceivable that shared aspects related to song contour in humans and songbirds may also be convergently coevolved with vocal learning. Future comparisons analyzing song contour variation among different species of birds, including vocal non-learners (e.g., common loons, *Gavia immer*) and in different modes of vocalization in humans (e.g., ingressive throat-rasping in Inuit/Ainu; Nattiez, 1999; Savage, Matsumae, et al., 2015) may help to distinguish between these or other potential alternatives, such as the adoption of learned "templates" (Ammirante & Russo, 2015).

Surprisingly, our analysis of contour in human songs around the world was more similar to Tierney et al.'s analysis of bird song than their analysis of the Essen folksong database. This difference might conceivably merely reflect the difference between notated scores vs. audio recordings. If this were the case, one would expect this to manifest in a more jagged appearance of the Essen contour reflecting jumps between discrete pitches vs. a smoother profile of the *Garland* contour based on continuous audio recordings, but in fact if anything the *Garland* contour appears more jagged (due to its smaller sample size). Thus, it seems that a more plausible interpretation of the difference might be that the Essen database is not the most representative sample upon which to base cross-cultural generalizations.

Although this study shows that the basic tendency towards descending/arched contours found in the Essen database does actually appear to generalize cross-culturally, more cross-cultural and cross-species research is warranted to determine to what degree other findings from music psychology are universal vs. specific to Western music (Patel & Demorest, 2013; Savage et al., 2015; Savage & Brown, 2013). Other candidates for future study include other features proposed to be explained by the motor constraint

hypothesis (e.g., small melodic intervals, phrase-final lengthening; Tierney et al., 2011), as well as features that have been proposed to be more promising candidates for biological specializations (e.g., beat synchronization; Patel, 2014). The *Garland* audio recordings provide a highly diverse but relatively small sample, so we hope that future studies will be able to test and extend the conclusions of ourselves and other corpus studies (most of which are primarily limited to Western classical/popular music, e.g., Huron, 2006; Temperley & VanHandel, 2013) using larger cross-cultural samples (e.g., Lomax, 1968) as they are made publicly available and new samples are developed.

Our successful use of automated audio analysis to compare relatively small samples of human and bird songs suggest promise for broader applications of "computational ethnomusicology" (Tzanetakis, Kapur, Schloss, & Wright, 2007) to larger samples incorporating additional musical features. On a cross-cultural level, this kind of rapid and objective audio analysis may make it feasible to re-attempt massive global analyses on the scale of Alan Lomax's (1968) Cantometrics Project and beyond. On a cross-species level, it likewise provides promise for empirically answering questions about issues such as parallels between human and bird songs (Araya-Salas, 2012; Bregman, Patel, & Gentner, 2016; Doolittle, Gingras, Endres, & Fitch, 2014). In this context, our findings provide yet another example of surprising similarities between human and animal vocalizations, highlighting the need for additional cross-cultural, cross-domain (e.g., music vs. language), and cross-species comparison to understand what is unique about human music.

5. *Author note*

We thank Emily Merritt for providing the manual phrase boundary annotations and Steven Brown, Bruno Gingras, and two anonymous reviewers for comments on an earlier version of this paper. *Correspondence concerning this article should be addressed to* Patrick E. Savage, Tokyo University of the Arts, Department of Musicology, Ueno-Kouen 12-8, Taito-ku, Tokyo 110-8714, Japan. Email: patsavagenz@gmail.com

6. References

- Ammirante, P., & Russo, F. A. (2015). Low-skip bias: The distribution of skips across the pitch ranges of vocal and instrumental melodies is vocally constrained. *Music Perception, 32*(4), 355–363.
- Araya-Salas, M. (2012). Is birdsong music? Evaluating harmonic intervals in songs of a Neotropical songbird. *Animal Behaviour, 84*(2), 309–313.
- Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences of the United States of America, 113*(6), 1666–1671.
- Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music, 41*(2), 229–248.
- Doolittle, E. L., Gingras, B., Endres, D. M., & Fitch, W. T. (2014). Overtone-based pitch selection in hermit thrush song: Unexpected convergence with scale construction in human music. *Proceedings of the National Academy of Sciences of the United States of America, 111*(46), 16616–16621. <http://doi.org/10.1073/pnas.1406023111>
- Elemans, C. P. H., Rasmussen, J. H., Herbst, C. T., Düring, D. N., Zollinger, S. A., Brumm, H., et al. (2015). Universal mechanisms of sound production and control in

- birds and mammals. *Nature communications*, 6: 8978. DOI: 10.1038/ncomms9978
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33, 61–135.
- Honing, H., Cate, C. ten, Peretz, I., & Trehub, S. E. (Eds.). (2015). Biology, cognition and origins of musicality [Special Issue]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1664).
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge: MIT Press.
- Lomax, A. (Ed.). (1968). *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.
- Nettl, B., Stone, R., Porter, J., & Rice, T. (Eds.). (1998). *The Garland encyclopedia of world music [10 volumes; 9 CDs]*. New York: Garland Pub.
- Patel, A. D. (2008). *Music, language and the brain*. Oxford: Oxford University Press.
- Patel, A.D. (2010). Music, biological evolution, and the brain. . In: M. Bailar, (Ed.). *Emerging Disciplines*. Houston, TX: Rice University Press (pp. 91–144).
- Patel, A. D. (2014). The evolutionary biology of musical rhythm: Was Darwin wrong? *PLoS Biology*, 12(3), e1001821. <http://doi.org/10.1371/journal.pbio.1001821>
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Savage, P. E., & Brown, S. (2013). Toward a new comparative musicology. *Analytical Approaches to World Music*, 2(2), 148–197.
- Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal

the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, 112(29), 8987–8992.

Savage, P. E., Matsumae, H., Oota, H., Stoneking, M., Currie, T. E., Tajima, A., ...

Brown, S. (2015). How “circumpolar” is Ainu music? Musical and genetic perspectives on the history of the Japanese archipelago. *Ethnomusicology Forum*, 24(3), 443–467. <http://doi.org/10.1080/17411912.2015.1084236>

Selfridge-Field, E. (1995). *The Essen musical data package*. Menlo Park, CA: Center for Computer Assisted Research in the Humanities (CCARH).

Suthers, R. A., Goller, F., & Pytte, C. (1999). The neuromuscular control of birdsong. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1385), 927–939.

Temperley, D., & VanHandel, L. (2013). Introduction to the special issues on corpus methods. *Music Perception*, 31(1), 1–3.

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

Tzanetakis, G., Kapur, A., Schloss, W. A., & Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2), 1–24.

Wallin, N. L., Merker, B., & Brown, S. (Eds.). (2000). *The origins of music*. Cambridge: MIT Press.

Tables

TABLE 1. Mean values (standard deviations in brackets) for pitch, melodic range, and duration of human and bird song contours. Note that the unit of analysis is based on breathing boundaries, corresponding to individual phrases for human songs and individual notes for bird songs. Values that could not be calculated from the Essen notations are marked as "NA".

	Pitch	Range	Duration
Mean values (SD) for:	(Hz)	(semitones)	(seconds)
Human (all Garland)	226 (135)	8.7 (3.6)	5.3 (3.4)
Human (Garland non-Western)	213 (134)	7.4 (3.3)	6.2 (3.8)
Human (Garland Western)	242 (134)	9.8 (3.4)	4.5 (2.8)
Bird	3,730 (1,584)	4.9 (4.1)	0.08 (0.07)
Human (Essen)	NA	8.9 (3.7)	NA

Figure captions

FIGURE 1. A map of the 35 audio recordings analyzed from the *Garland Encyclopedia of World Music*. “Western” songs from Europe or by speakers of European languages (e.g., English, Spanish, Portuguese) are colored light gray, “non-Western” songs are colored dark gray.

FIGURE 2. Example contour analysis using the first four phrases of one of the songs analyzed (Chinese solo *fanbei* chant, ID no. 190). The top panel shows the four phrases as spectrograms, while the lower panel shows the same phrases as normalized pitch contours, with the y-axis showing pitches as semitones above or below the mean and the x-axis showing normalized time at 50 equally spaced time points from the beginning to the end of the phrase.

FIGURE 3. Comparison of song contours from the *Garland Encyclopedia of World Music* with those from the bird song database and Essen musical database reported by Tierney et al. (2011). Average pitch–time contours were generated by sampling each normalized pitch contour at 50 equally spaced time points, then averaging across contours. Shaded lines indicate SE for each time point (too small to see for the Essen sample due to the large sample size). The *Garland* contours are significantly more correlated with the bird song contours ($r = .95$) than with the Essen contours ($r = .88$; $p < .01$).

Figures

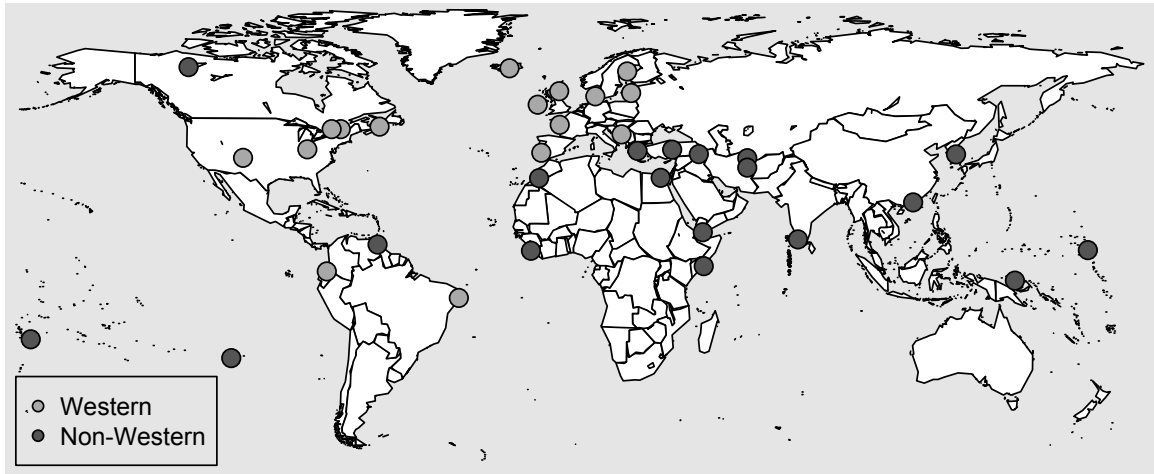


FIGURE 1

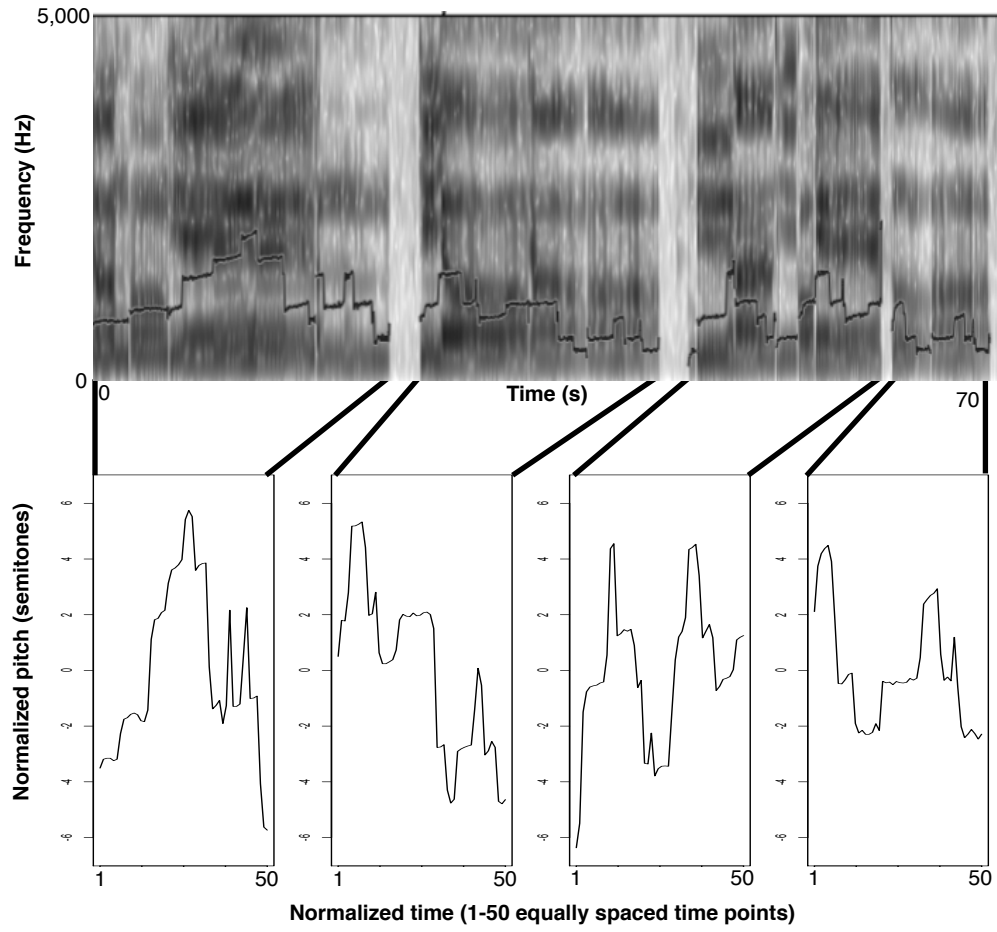


FIGURE 2

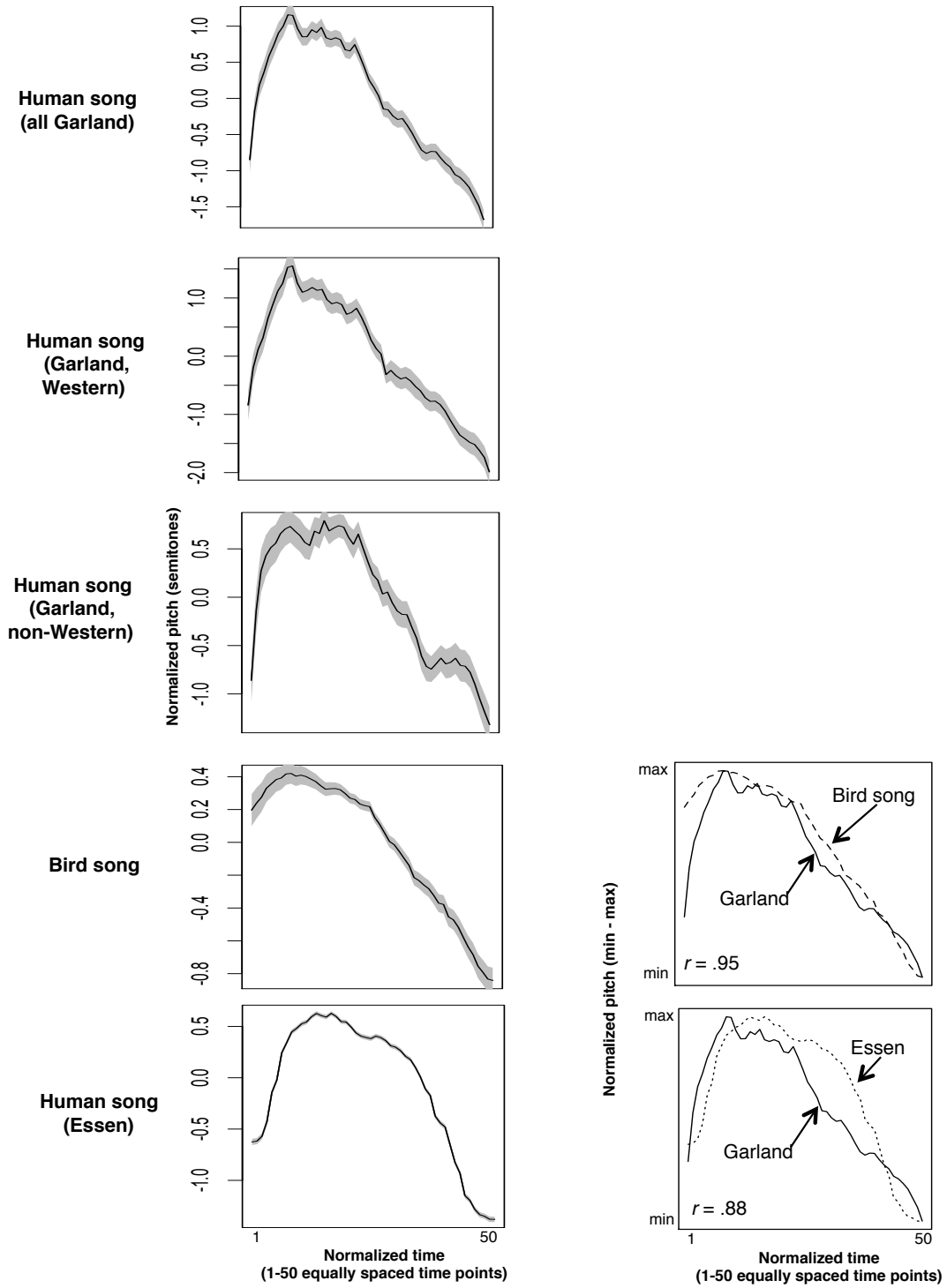


FIGURE 3