

Interval Size and Phrase Position: A Comparison between German and Chinese Folksongs

DANIEL SHANAHAN
Ohio State University

DAVID HURON[1]
Ohio State University

ABSTRACT: It is well known that the pitch of the voice tends to decline over the course of a spoken utterance. Ladd (2008) showed that there is also a tendency for the pitch range of spoken utterances to shrink as the pitch of the voice declines. Motivated by this work, two studies are reported that test for the existence of “late phrase compression” in music where the interval size tends to decline toward the end of a phrase. A study of 39,863 phrases from notated Germanic folksongs shows the predicted decline in interval size. However, a second study of 10,985 phrases from Chinese folksongs shows a reverse relationship. In fact, the interval behaviors in Chinese and Germanic folksongs provide marked contrasts: Chinese phrases are dominated by relatively large intervals, but begin with small intervals and end with medium-small intervals. Germanic phrases are dominated by relatively medium intervals, but begin with large intervals and end with small intervals. In short, late phrase interval compression is not evident cross-culturally.

KEYWORDS: *Melody, culture, declination, late-phrase compression, contour, prosody*

PATEL (2008) has documented many ways in which music and language share common features. The pitch contour of speech is one of the central elements of prosody. In speech, there is a strong tendency for utterances to exhibit a downward pitch trajectory – referred to as *declination* (‘t Hart & Cohen, 1973; Vassière, 1983). ‘t Hart, Collier, and Cohen (1990) suggested that the proximal cause of speech declination is the reduction in sub-glottal air pressure as the volume of air in the lungs is exhausted over the course of an utterance. Ethnomusicologist Curt Sachs (1962) has noted that many musical cultures employ “tumbling” phrases in their vocal music. More formally, Huron (2006) showed that pitch contours in a wide sample of melodic phrases also commonly exhibit pitch declination, however, the effect size in music is only about one-tenth of that evident in speech. This reduced effect size may arise due to the musical tendency to produce specific stable pitches which necessitates more sustained air support.

Another common pitch contour pattern is the so-called *melodic arch* where phrases tend to ascend at the beginning of the phrase and descend in the latter half of the phrase. This arch-shaped melodic tendency is evident in Figure 1 which reproduces an illustration from Huron (1996) showing the average pitch contour for 6,364 seven-note phrases from European folksongs.

Ladd (2008) has observed another general prosodic tendency where the pitch range of spoken utterances tends to shrink as the utterance progresses.[2] To our knowledge, no research has been carried out to determine whether this ubiquitous speech-related pattern is also apparent in musical melodies. In this article, we report two studies whose purpose is to test this relationship. To anticipate our results, we will see evidence of *late-phrase compression* in a sample of Western music, however, the reverse relationship is evident in a sample of Chinese music. We conclude that there is no evidence of late-phrase compression as a general principle of melodic organization, and that cultural patterns play a more important role.

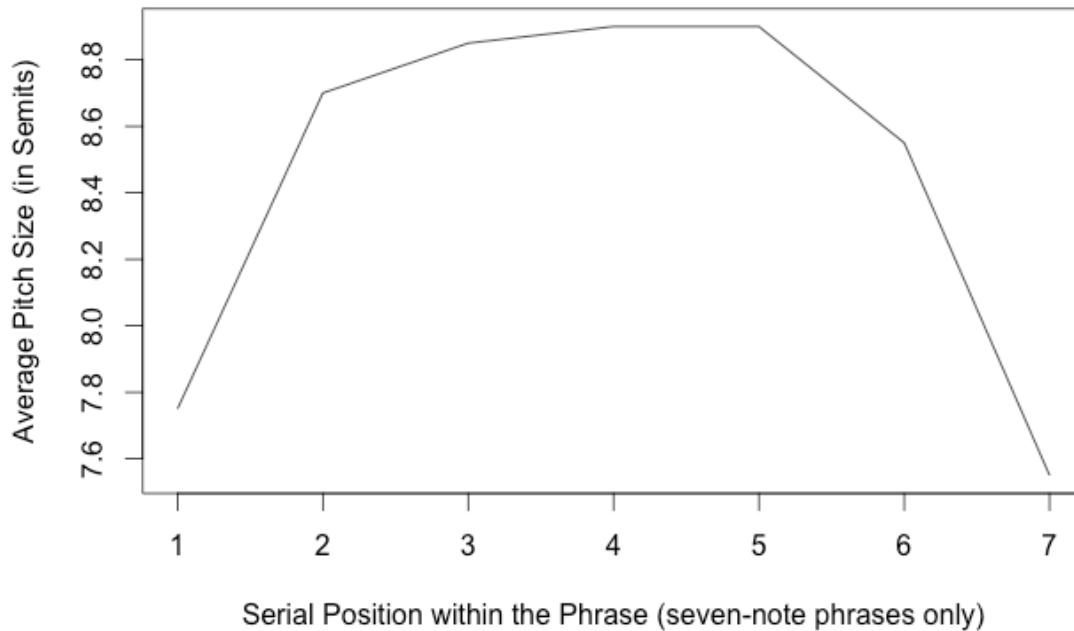


Fig. 1. Average contour of seven-note phrases in European folksongs; the graph depicts average pitch height measured in semitones above middle C. After Huron (1996).

HYPOTHESIS

Formally, we may state our hypothesis as follows:

H1. Melodic phrases exhibit a general tendency where the variability in pitch movement declines over the duration of the phrase.

In brief, we test this hypothesis using two contrasting samples of musical melodies.

STUDY 1: PITCH REGRESSION IN GERMANIC FOLKSONGS

In order to test our hypothesis we need to operationalize the various terms. In particular, we need to consider “what is a phrase?” and “what do we mean by pitch variability?”

In speech research, “phrases” are usually operationalized as speech utterances bounded by points of speaker inhalation. At least in the case of Western music theory, the breath also figures as a prominent indicator of a phrase boundary. However, since much music employs instruments that do not involve breathing, the phrase concept has been generalized beyond the inter-breath unit considered appropriate for speech or song. In studying musical phrases, vocal music provides an especially appropriate source since fewer presumptions are made concerning the meaning of “phrase.” In contrast to art-song (where an expert vocalist may be challenged to sustain especially long musical lines), folksongs arguably exhibit a closer association between the nominal phrasing and actual patterns of breathing.

Accordingly, we chose to examine phrasing in an existing database of Germanic folksongs – specifically the Essen Folksong Collection (Schaffrath, 1995). Each of the 6,251 folksongs in the Essen Folksong Collection includes pitch and duration information, meter signatures, barlines, rests, and explicit phrase markings. The database was created from a number of notated sources assembled by Helmuth Schaffrath. Many of these sources are from materials collected by ethnomusicologists or by folklorists.

The provenance of the phrase markings in the Essen database is unknown. When transcribing recorded songs, ethnomusicologists and folklorists commonly, though not always, transcribe phrases at breath points. However, we do not know whether all of the phrase markings included in the Essen Collection originated in transcriptions, if so, whether the folklorist used breath points as cues to phrase markings, or whether Schaffrath and his colleagues added explicit phrase markings to notated sources that did not contain such markings. If phrase markings were added by Schaffrath to the database, then they would have been introduced without reference to the sound recordings in which the breathing might have guided the coding. Consequently, we cannot discount the possibility that our results may be confounded by whatever criteria or mental models the researchers used when coding the folksongs. For the purposes of this study, we will tentatively assume that the phrase indications encoded by Schaffrath and his colleagues exhibit no systematic biases that might unduly influence our measurements.

With regard to pitch variability, once again, we must assume that the notated pitches in the Essen database represent reasonable approximations of the actual or intended vocal action. The more pertinent question is how we characterize pitch *variability*. In the case of Ladd (2008), the variability is expressed in terms of range. In contrast to speech, music makes greater use of relatively fixed or categorical pitches. A convenient way to characterize pitch movement is through the common musical concept of the melodic interval. Between any two pitches, one might, for example, calculate the number of semitones (semit). Accordingly, we can use the *average interval size* (in semits) as an index for the pitch variability.

In short, we operationalize our hypothesis as predicting that the average interval size will decrease over the course of the phrase.

PROCEDURE

All of the sampled materials are available in the Humdrum “kern” format (see <http://kernscores.org>). We extracted all of the phrases from a sample of 5,332 folksongs using the Humdrum Toolkit (Huron, 1995, 2002). In total, some 30,846 phrases were extracted, ranging in length from 1 to 29 notes.

Some melodic phrases include notated rests. The presence of these rests raises some questions. Rests are often points where a singer takes a breath. Especially for longer rests, the very presence of a rest in the middle of a notated phrase might raise doubts about whether the phrase is correctly coded. Rests also raise questions regarding the determination of melodic intervals. Pitch intervals are perceptually less salient when they are not part of a perceptual group. For example, musicians report that they are often oblivious to the pitch interval between the last note of one phrase and the first note of the next phrase. Since the presence of rests raises questions about the status of the phrase as well as the nature of the melodic interval, we decided to circumvent these potential pitfalls by excluding from consideration any nominal phrases containing a notated rest in the midst of the phrase. Given this criterion, the sample size was reduced by 567 phrases, from 30,846 phrases to 30,281 phrases. Four aberrant phrases in the database ended up containing just rests, which, when removed, resulting in a final tally of 30,277 phrases. Table 1 shows the number of phrases for each phrase length in this reduced sample.

No. of notes	No. of phrases	No. of notes	No. of phrases
1	7	15	93
2	135	16	37
3	570	17	32
4	992	18	12
5	3,912	19	9
6	5,614	20	7
7	7,321	21	5
8	4,019	22	4
9	3,307	23	3
10	1,929	24	1
11	1,283	25	1
12	500	29	1
13	361	30	0
14	122		

Table 1. Frequency of different phrase lengths in sampled German folksongs.

The majority of phrases (98.8%) were between 3 and 14 notes in length. One- or two-note phrases, or phrases containing more than 14 notes, are less representative of common phrase-related behavior. Consequently, we limited our analyses to phrases containing between 3 and 14 notes.

RESULTS

Figures 2 and 3 display the average interval size according to serial note position for phrases from 3 to 14 notes in length. Regression slopes are also plotted for each figure. As can be seen, with the exception of 6-note phrases, the regression slopes are all negative. At face value, the predominance of negative slopes is consistent with the hypothesis that pitch variability is reduced towards the ends of phrases.

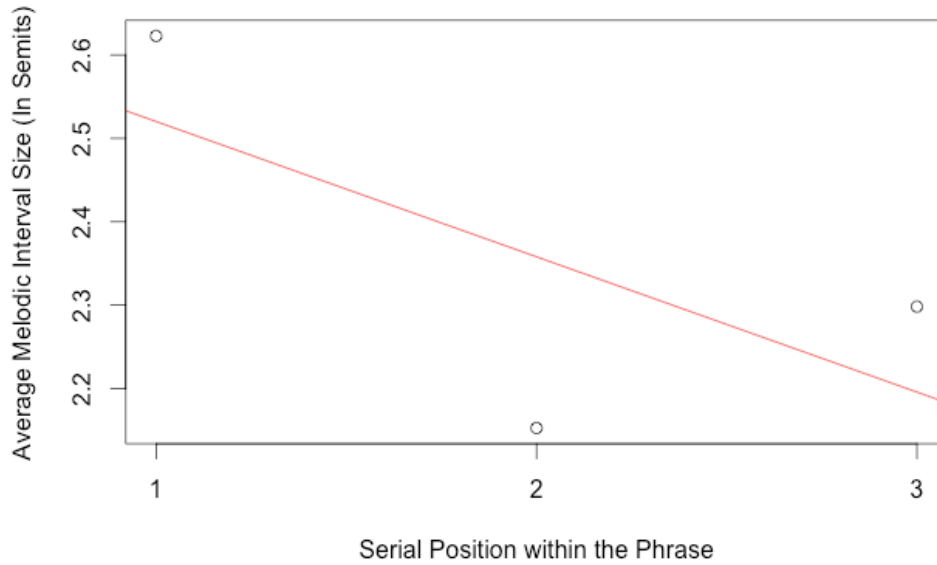
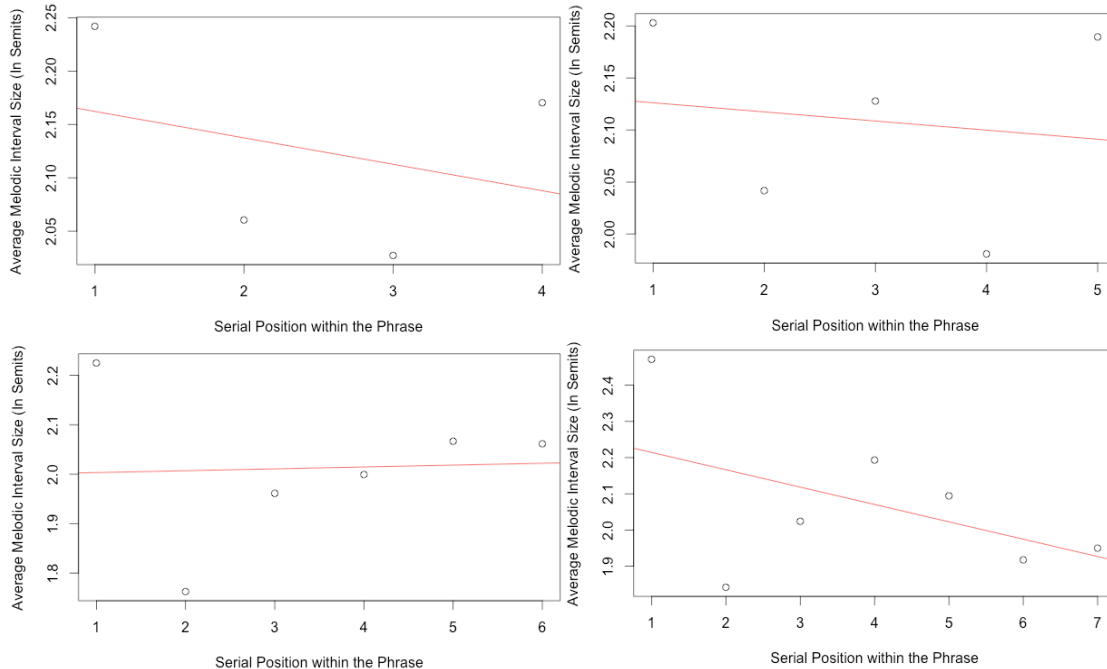


Fig. 2. Average interval size for 570 three-note phrases with accompanying regression slope.



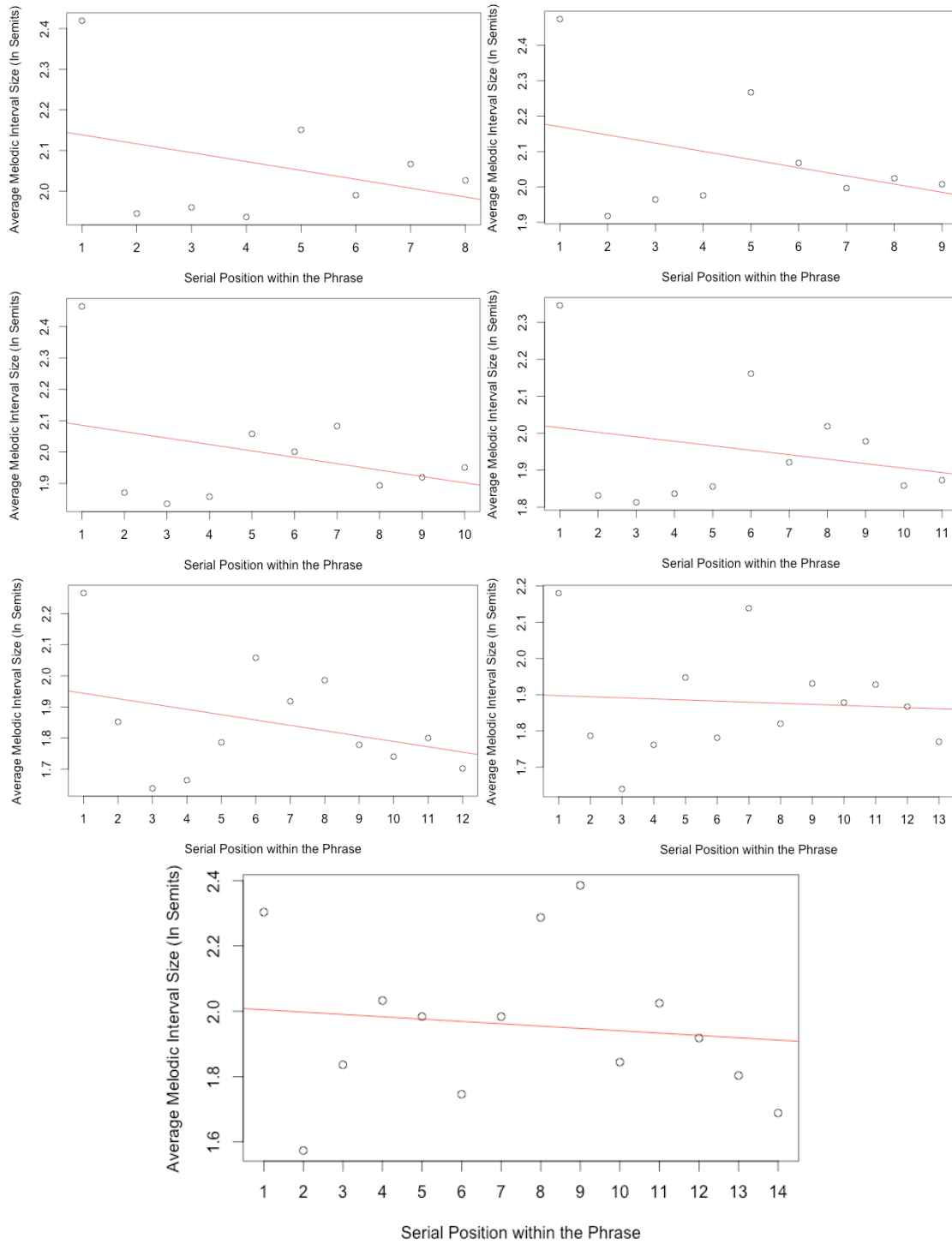


Fig. 3. Average interval size for phrases containing between 4 and 14 notes, with their accompanying regression slopes. Note the large initial interval in each panel. Discussion of this feature will be deferred until the Post-Hoc Observation section.

In order to test the significance of this apparent effect, we calculated the regression slopes for each phrase individually. Accordingly, we measured the rank order correlation for each of the 30,277 candidate phrases, comparing the average melodic interval size (in semits) versus serial position in the phrase.

Excluding correlation values of zero, 13,478 phrases exhibited positive slopes and 15,011 phrases exhibited negative slopes. As a conservative statistical test, we can treat each phrase as a single independent observation, and simply compare the ratios of positive to negative slopes. Although this difference amounts to just five percent, it remains statistically significant ($X^2 = 82.5$; $p < 0.00005$; one-tailed test). Hence, at face value, the results are consistent with the hypothesis.

A striking feature evident in Figure 3 is the predominance of a large initial interval at the beginning of each phrase, and perhaps a less striking tendency for the final interval to be small. The intervening intervals appear to show less of a consistent downward trend. As a consequence, the negative slopes may merely be an artifact of the first and last intervals. We will consider these and other patterns in more detail in the section below, entitled Post-Hoc Observations.

STUDY 2: PITCH REGRESSION IN CHINESE FOLKSONGS

A number of potential problems might be identified regarding Study 1. The foremost concern is that the sample consists exclusively of European folksongs, specifically of Germanic origin. It would be appropriate to test the same hypothesis using music of non-Western origin. Accordingly, for Study 2 we made use of a convenience sample of Chinese folksongs, also assembled by Helmut Schaffrath and his colleagues. Specifically, Schaffrath's database consists of 2,250 traditional Chinese folksongs, primarily of Han origin. Once again, we have no knowledge regarding the provenance of the phrase markings.

PROCEDURE

All of the materials are available in the Humdrum "kern" format, and processing was carried out using the Humdrum Toolkit. The procedure was the same as for Study 1. Phrases containing mid-phrase rests were discarded.

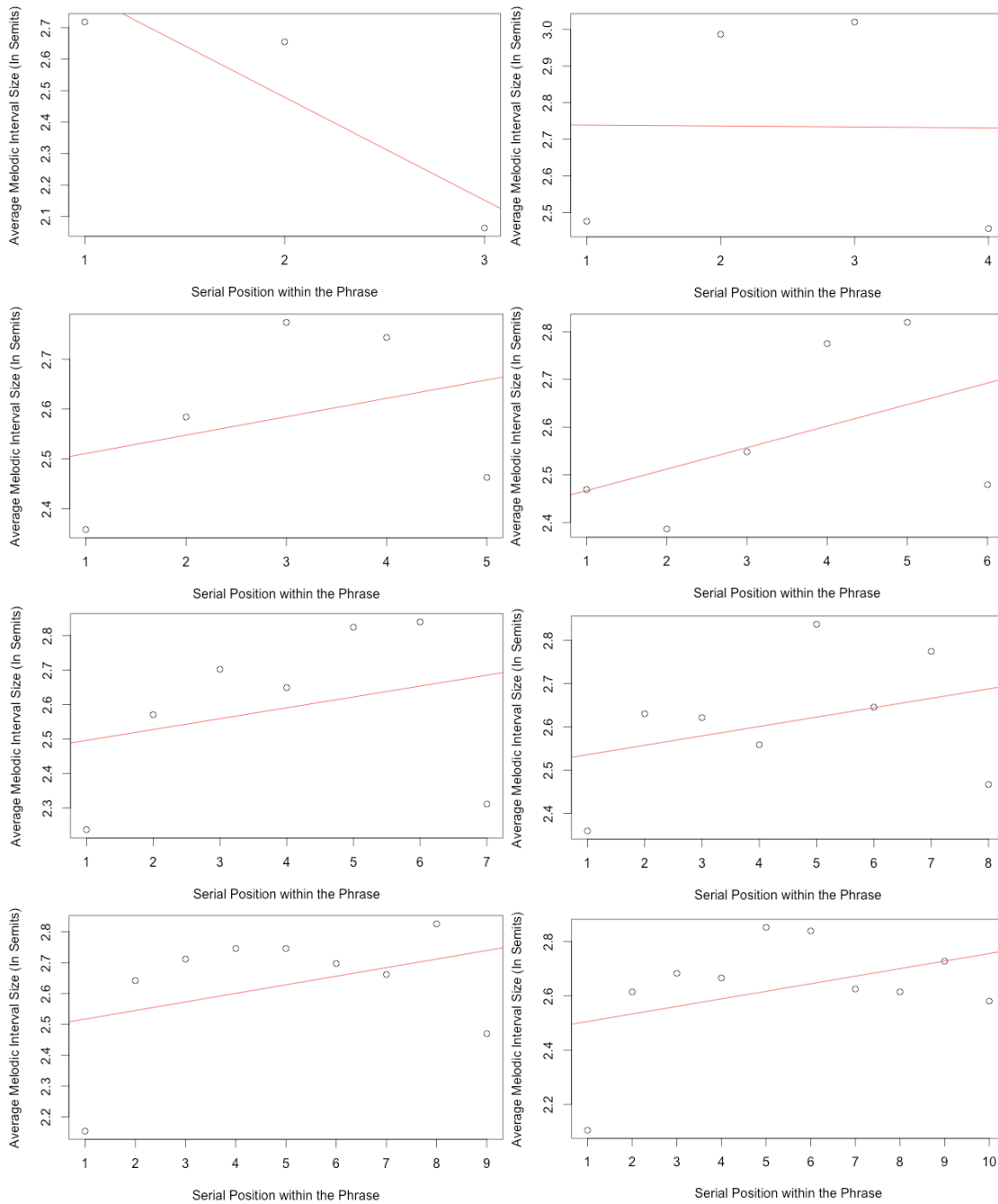
RESULTS

In total, some 11,577 phrases were extracted, ranging in length from 1 to 169 notes. Excluding phrases containing rests, the sample size was reduced by 592 phrases, to 10,985 phrases. In contrast to the Germanic folksongs, the Chinese folksong database contained a number of items in which the entire song was coded as a single phrase. Consequently, there were 39 nominal "phrases" between 31 and 169 notes in length. It is not known whether this variability reflects genuine characteristics of Chinese folk music, or whether these results are artifacts of the database coding. Table 2 shows the number of phrases for each phrase length from 1 to 30 notes in length.

No. of notes	No. of phrases	No. of notes	No. of phrases
1	30	16	425
2	90	17	390
3	142	18	318
4	298	19	256
5	363	20	194
6	582	21	128
7	715	22	104
8	784	23	73
9	902	24	58
10	1,006	25	32
11	1,023	26	18
12	931	27	16
13	806	28	11
14	640	29	9
15	580	30	2

Table 2. Frequency of different phrase lengths in sampled Chinese folksongs.

In the Germanic folksongs, 98.8% of the phrases were between 3 and 14 notes in length. For the Chinese folksongs only 74.6% of the phrases were between 3 and 14 notes in length. If phrases are considered between 3 and 30 notes in length, then 98.4% of the phrases are represented. Once again, separate analyses were carried out for phrases of different lengths. Figure 4 displays the average interval size according to note position for phrases from 3 to 20 notes in length (representing 95.4% of all phrases). Regression slopes are also shown. With the exception of 3- and 4-note phrases, the regression slopes are clearly positive – contrary to the hypothesis. Notice that the regression slopes may be unduly influenced by the initial interval, which, in Figure 4, are conspicuously small in size. Further discussion of this is deferred to the section entitled Post-Hoc Observations.



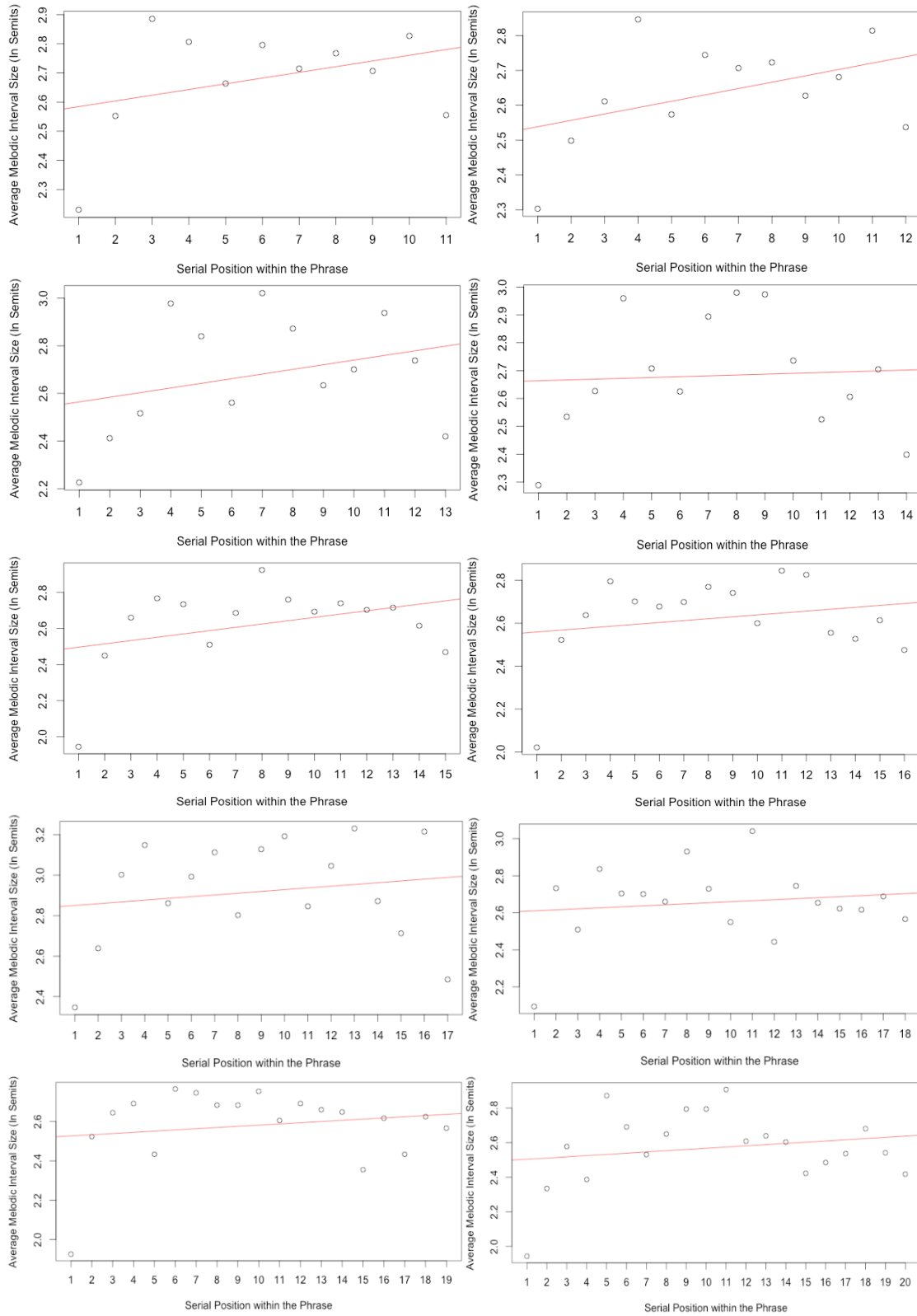


Fig. 4. Average interval size for phrases containing between 4 and 20 notes, with their accompanying regression slopes.

Once again, we calculated the regression slopes for each phrase individually. Excluding correlation values of zero, we counted 5,485 phrases exhibiting positive slopes and 4,583 phrases exhibiting negative slopes. Although this difference amounts to just six percent, it is statistically significant ($X^2 = 80.8$; $p < 0.0001$; two-tailed test), but in the opposite direction predicted by the hypothesis. Unlike the Germanic folksongs, the Chinese folksongs are more consistent with late-phrase *expansion* rather than late-phrase compression.

CONCLUSION

While the results of the first study are consistent with the experimental hypothesis, the results of the second study are wholly inconsistent with the hypothesis. In short, we cannot conclude that there is a broad musical tendency, analogous to the prosodic tendency for pitch variability to decline over the course of an utterance. Instead, the results highlight differences between the two samples. The natural tendency would be to conclude that these differences reflect cultural factors, however, one cannot rule out the possibility that the results are genre-based rather than culture-based. That is, we cannot necessarily assume that German folksongs are representative of Western music in general, or that Chinese folksongs are representative of Chinese music in general. Further studies would be necessary in order to support either claim. Nevertheless, the results strongly implicate culture as the principal source of the observed differences.

Although our results are not consistent with the hypothesis motivated by the speech prosody research, the negative results reported here should not necessarily be construed as suggesting that language plays no role in phrase-related melodic organization. It may very well remain the case that the observed differences between Germanic and Chinese folksongs relate in some way to linguistic factors. In order to address this matter, we carried out some further post-hoc analyses.

POST-HOC OBSERVATIONS

In examining the interval patterns in Germanic folksong phrases, there appears to be a consistent pattern. With the exception of 14-note phrases, the largest average interval is found at the beginning of the phrase. Interestingly, there is a strong tendency for the second interval to be among the smallest intervals in the phrase. The final interval appears to be roughly average in size.

In the case of the Chinese folksong phrases, there appears to be a very different pattern. With the exception of 3- and 4-note phrases, the smallest average interval is found at the beginning of the phrase. In most cases, the last interval of the phrase is the second smallest. For phrases between 6 and 9 notes in length, the largest average interval appears to occupy the penultimate position. However, for other phrase lengths, the largest average interval appears to occupy a position toward the center of the phrase.

There are a number of possible factors that might account for these differences. Mandarin is a tonal language, which makes use of four tones: a high tone, a rising tone, a falling-rising tone, and a falling tone. Notice that these tonal patterns are asymmetrical. For example, there is no low tone to match the high tone and there is no rising-falling tone to match the falling-rising tone. Moreover, we have no evidence to suggest that the number of rising tones is equivalent to the number of falling tones. The observed differences between the German and Mandarin interval patterns might be attributable to some aspect of tone use in Mandarin.

Another possibility is that the differences in interval behavior relate to common stress patterns in German and Mandarin. Indo-European languages tend to exhibit a so-called *proclitic* rhythmic organization where a host morpheme is preceded by a weakly stressed grammatical element. Examples of such weak-to-strong groupings include *to bed*, *der Apfel*, and *du bois*. Dynamic and agogic stresses are often echoed in pitch related treatments. For example, in an extensive study of melodic accent, Huron and Royal (1996) observed that, after Thomassen's model of melodic accent, interval size is the second strongest correlate of metrical stress (see also Thomassen, 1982, 1983). That is, for a wide sample of (Western) musics, there is a correlation between large interval size and position in the metric hierarchy. Large intervals tend to be accompanied by weak-to-strong rhythmic movements.

The tendency for large average intervals to occur at the beginnings of phrases in Germanic folksongs is consistent with an anacrusis or "pick up" where the rhythm follows a weak-to-strong pattern. In Chinese folk music, the relatively small size of the initial average interval in the phrase is consistent with a strong-to-weak rhythmic pattern where the small interval is symptomatic of a weak stress accorded the

second note. This could be related to the phenomenon of *duì* character- or syllable-pairing in Mandarin, consistent with *enclitic* treatment, where the host morpheme is followed by a weaker syllable (Shen, 1990). Such contrasts have already been observed by Iversen, Patel and Ohgushi (2006), who demonstrated how proclitic and enclitic linguistic patterns influence the perception of musical grouping in English and Japanese listeners.

In order to test this possibility, we carried out a post-hoc analysis of the metric relationships between the first two notes in both the Germanic and Chinese samples. Specifically, we assumed the traditional hierarchy for common meter signatures like 3/4 and 4/4, where the first beat in the measure is the strongest, with successive beats and sub-beats assuming progressively less stressed values. For example, in 2/4 meter, we might characterize the stress rankings for 8 sixteenth notes as: 1, 4, 3, 4, 2, 4, 3, 4 – indicating that the highest stress accrues to the first note of the measure, the second highest to the second beat, and so on. Using this coding method, we determined the stress values for the first and second notes in subsamples for both the German and Chinese folksongs. Using a sample of 299 German folksongs, 241 were found to begin with a weak-to-strong pattern in the first two notes (58 strong-to-weak). For a sample of 260 Chinese folksongs, fully 259 begin with a strong-to-weak pattern, with only 1 exhibiting a weak-to-strong opening.

Having made the proclitic/enclitic distinction, it is possible to interpret a general trend in the interval behavior of both the Germanic and Chinese folksongs. Recall that melodic accent is strongly correlated with large pitch intervals; that is, large intervals tend to occur in weak-to-strong rhythmic contexts, whereas small intervals tend to occur in strong-to-weak contexts (Huron & Royal, 1996). As a *proclitic* language, we might predict that Germanic folksongs would tend to begin and end with weak-strong syllable pairs. As a consequence, we should see evidence of large initial and final intervals. As an *enclitic* language, we might predict that Chinese folksongs would tend to begin and end with strong-to-weak syllable pairs. As a consequence, we should see evidence of small initial and final intervals. At the same time, melodies exhibit a tendency toward arch-shape contours, with the highest pitches occurring near the center or slightly beyond the center of the phrase (Huron, 1996). Pitches that are most distant from the center of the tessitura tend to be approached by the largest intervals (von Hippel, 2000; von Hippel & Huron, 2000). Accordingly, we would expect a large interval to occur near the center or beyond the center of the phrase. Given this scenario, we would predict that Germanic folksongs would tend to begin and end with large intervals, with one or more additional large intervals near the center of the phrase. Similarly, we would predict that Chinese folksongs would tend to begin and end with small intervals, with one or more large intervals near the center of the phrase. Visual inspection of figures 2-4 appears to be consistent with this post-hoc interpretation.

The above interpretation notwithstanding, the initial interval sizes appear to be strongly correlated with pervasive linguistic rhythms found in the two languages associated with the respective musical repertoires (see also Patel, 2008). While the current study finds no support for late-phrase compression, it nevertheless points to the possible influence of language in shaping melodic organization. Future research might consider using cluster analysis or other multi-variate exploratory methods to see if there are specific melodic patterns that dominate for different phrase lengths and different cultures.

NOTES

[1] Please address all correspondence to: David Huron, School of Music, 1866 College Rd., Columbus, OH, 43210, U.S.A.

[2] Ladd (2008) expresses this tendency in slightly different terms than given here. Ladd noted the tendency of the pitch range to shrink as the pitch of the voice declines. He presumed that all utterances exhibit a tendency toward pitch declination, and so lower pitch is necessarily correlated with later position in the phrase. At least in the case of Western melodies, declination is less pronounced than in speech, and arch-shaped phrases predominant. If Ladd's original formulation is taken literally, then one would expect interval size to be proportional to pitch height, with larger intervals involving higher overall pitches. We do not test this conjecture in this study.

REFERENCES

- Huron, D. (1995). *The Humdrum Toolkit: Reference Manual*. Menlo Park, California: Center for Computer Assisted Research in the Humanities.
- Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology*, Vol. 10, pp. 3-23.
- Huron, D. (2002). Music information processing using the Humdrum Toolkit: Concepts, examples, and lessons. *Computer Music Journal*, Vol. 26, No. 2, pp. 15-30.
- Huron, D., & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception*, Vol. 13, No. 4, pp. 489-516.
- Iversen, J.R., Patel, A.D., & Ohgushi, K. (2006). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, Vol. 124, No. 4, pp. 2263-2271.
- Ladd, D.R. (2008). *Intonational Phonology*, 2nd edition. Cambridge: Cambridge University Press.
- Patel, A.D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Sachs, C. (1962). *The Wellsprings of Music*. The Hague: Martinus Nijhoff.
- Schaffrath, H. (1995). *The Essen Folksong Collection*. D. Huron (Ed.), Stanford, California: Center for Computer Assisted Research in the Humanities.
- Shen, X.N. (1990). *The Prosody of Mandarin Chinese*. Berkeley: University of California Press.
- 't Hart, J., & Cohen, A. (1973). Intonation by rule: A perceptual quest. *Journal of Phonetics*, Vol. 1, pp. 309-321.
- 't Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation; An Experimental-phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.
- Thomassen, J. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America*, Vol. 71, pp. 1596-1605.
- Thomassen, J. (1983). Erratum. *Journal of the Acoustical Society of America*, Vol. 73, No. 1, p. 373.
- Vassière, J. (1983). Language - independent prosodic features. In: A. Cutler & D.R. Ladd (Eds.), *Prosody: Models and Measurements*. Berlin: Springer, pp. 53-66.
- von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception*, Vol. 17, No. 3, pp. 315-327.
- von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, Vol. 18, No.1, pp. 59-85.