Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Prosodic Stress, Interval Size and Phrase Position: A Cross-Cultural Contrast

Daniel Shanahan,[1] David Huron[2]

Ohio State University
[1]shanahan.37@osu.edu, [2]huron.1@osu.edu

ABSTRACT

Two studies were carried out in order to test the existence of "late phrase compression" in music where the interval size tends to decline toward the end of a phrase. A sample of phrases from notated Germanic folksongs shows the predicted decline in interval size. However, a sample of phrases from Chinese folksongs shows a reverse relationship. In short, late phrase interval compression is not evident cross-culturally.

I. INTRODUCTION

In research pertaining to speech prosody, the existence of a downward trajectory of pitch– commonly referred to as *declination* –has been explored at length (see (t'Hart & Cohen, 1973; Vassière, 1983, Ladd, 2008). Sachs (1962) has similarly noted that many musical cultures employ "tumbling" phrases in their vocal music. A related general prosodic tendency, in which the the pitch range of spoken utterances tends to shrink as the utterance progresses. (Ladd, 2008) To our knowledge, there has been no work carried out that investigates whether this ubiquitous speech-related pattern is also apparent in musical melodies.

II. HYPOTHESIS

Formally, we may state our hypothesis as follows:

H1. Melodic phrases tend to decline in pitch movement and pitch variability over the course of the phrase.

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

III. STUDY #1: PITCH REGRESSION IN GERMANIC FOLKSONGS

Before continuing, it might first be prudent to operationalize the definitions of both phrase and pitch variability. In speech research, "phrases" are usually defined as 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 a great deal of 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 origin of the notated phrase markings in the Essen database is unknown. Consequently, we cannot discount the possibility that our results may be somewhat 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. For this study, we have used 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.

A. Procedure

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

The occasional presence of rests raised some early questions. 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. 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.

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

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

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.

B. Results

Figures 1 and 2 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.

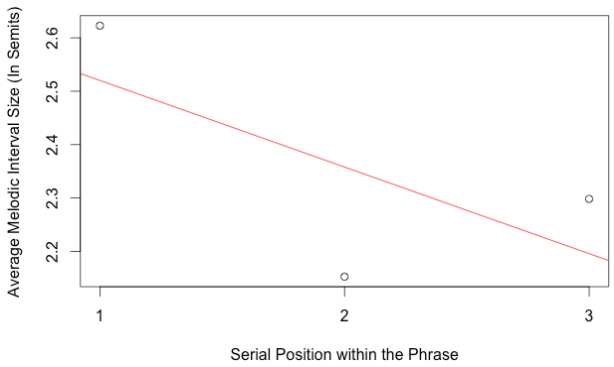
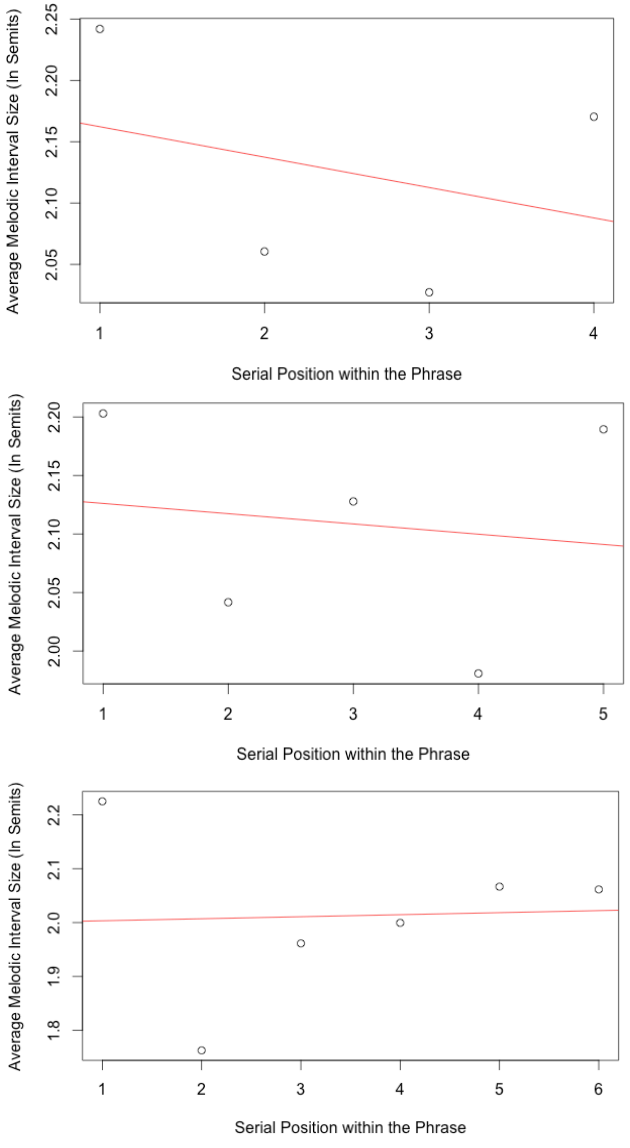


Figure 1. Average interval size for 570 three-note phrases with accompanying regression slope.



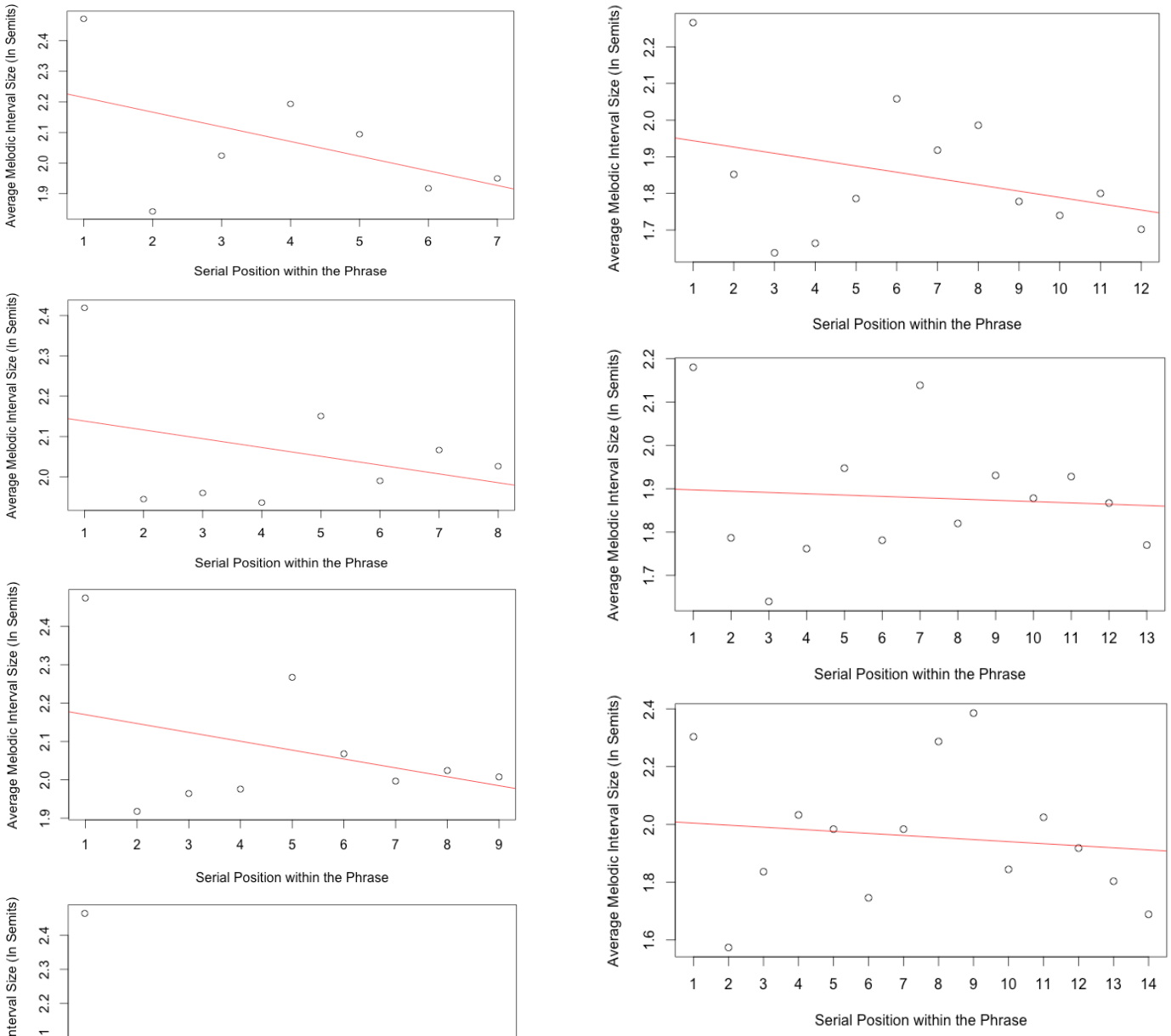


Figure 2. Average interval size for phrases containing between 4 and 14 notes, with their accompanying regression slopes.

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 (χ^2=82.5; p<0.00005; one-tailed test). Hence the results are consistent with the hypothesis.

IV. STUDY #2: PITCH REGRESSION IN CHINESE FOLKSONGS

A number of potential problems might be identified regarding Study 1. Most notably, 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.

A. Procedure

The procedure for the second experiment was the same as for Study #1. Phrases containing mid-phrase rests were discarded.

B. Results

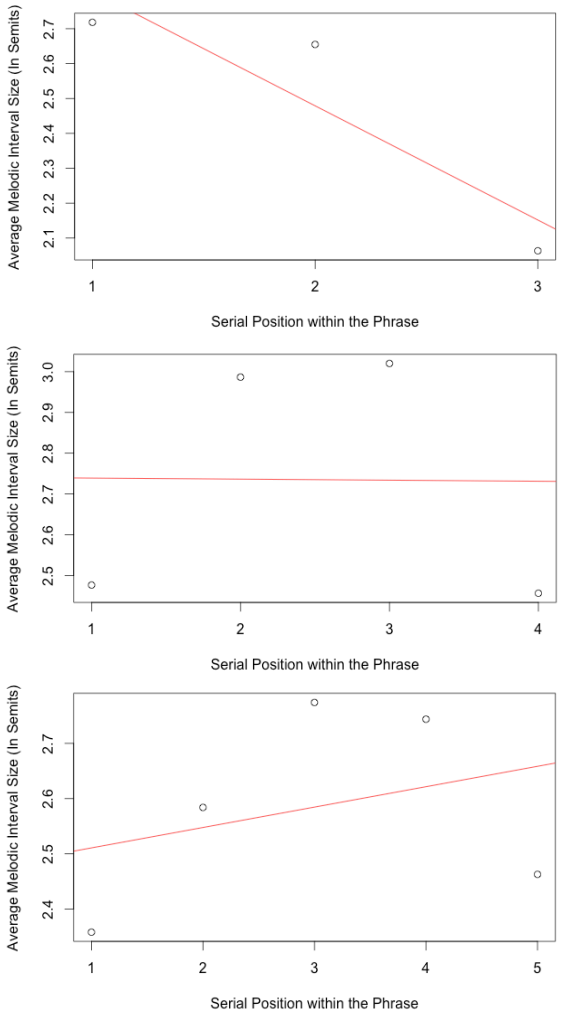
In total, some 11,577 phrases were extracted, ranging in length from 1 to 169 notes. Excluding phrases containing rests, the sample size was reduced by 592 phrases, to 10,985 phrases. In contrast to the Germanic folksongs, the Chinese folksong database contained a number of items in which the entire 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.

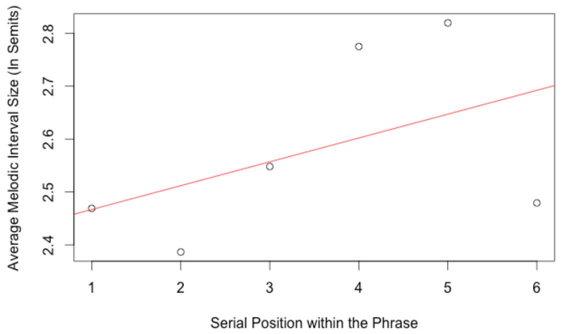
Table 2: Frequency of different phrase lengths in sampled Chinese folksongs.

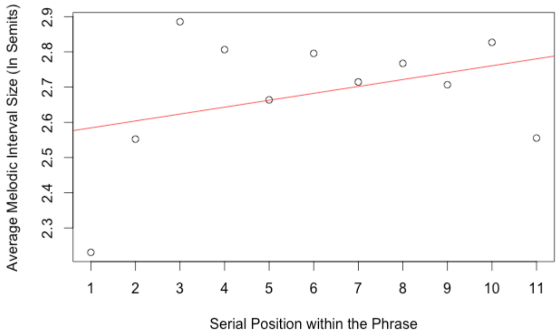
No. of notes	No. of phrases
1	30
2	90
3	142
4	298
5	363
6	582
7	715
8	784
9	902
10	1,006
11	1,023
12	931
13	806
14	640
15	580
16	425
17	390
18	318
19	256
20	194
21	128
22	104
23	73
24	58
25	32
26	18
27	16
28	11
29	9
30	2

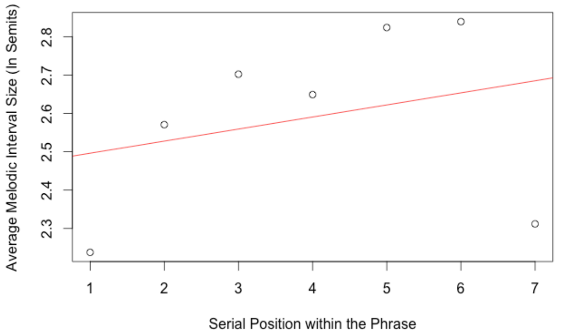
In the Germanic folksongs, 98.8% of the phrases were between 3 and 14 notes in length. For the Chinese folksongs only 74.6% of the phrases were between 3 and 14 notes in length. If phrases are considered between 3 and 30 notes in length, then 98.4% of the phrases are represented. Figure 3 displays the 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.

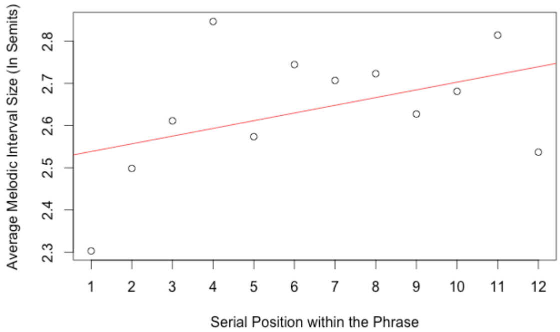
Figure 3. Average interval size for phrases containing between 4 and 20 notes, with their accompanying regression slopes.

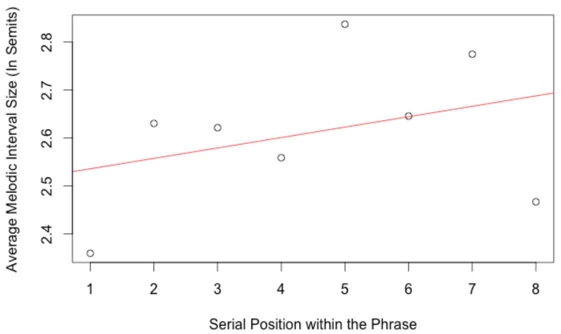


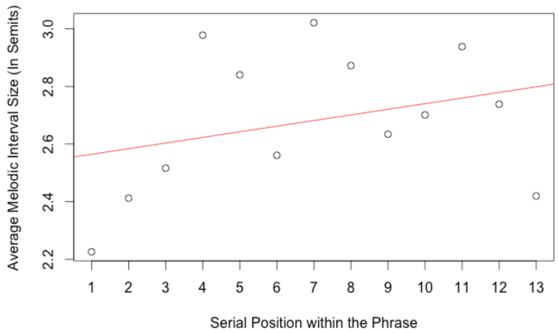

Average Melodic Interval Size (In Semits)
2.4
2.5
2.6
2.7
2.8
1
2
3
4
5
6
Serial Position within the Phrase

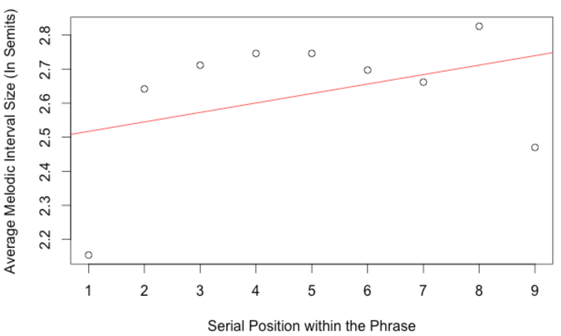

Average Melodic Interval Size (In Semits)
2.3
2.4
2.5
2.6
2.7
2.8
2.9
1
2
3
4
5
6
7
8
9
10
11
Serial Position within the Phrase

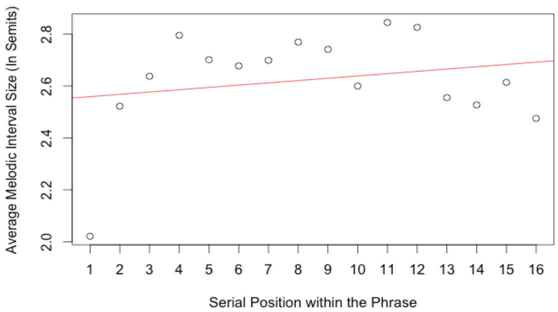

Average Melodic Interval Size (In Semits)
2.3
2.4
2.5
2.6
2.7
2.8
1
2
3
4
5
6
7
Serial Position within the Phrase

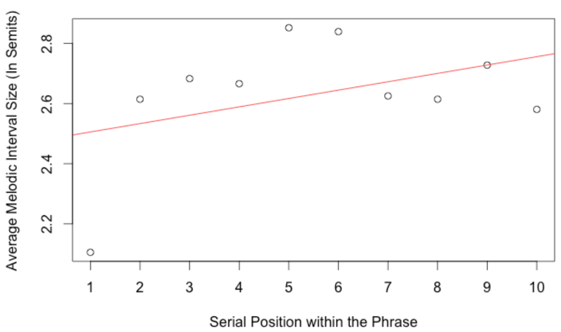

Average Melodic Interval Size (In Semits)
2.3
2.4
2.5
2.6
2.7
2.8
1
2
3
4
5
6
7
8
9
10
11
12
Serial Position within the Phrase

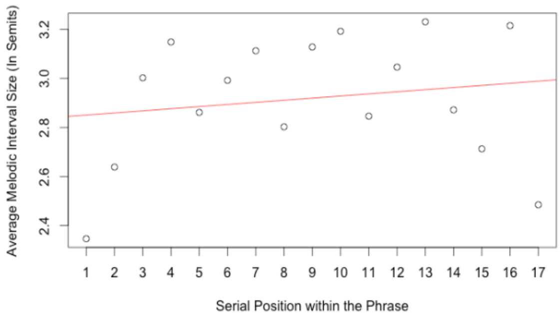

Average Melodic Interval Size (In Semits)
2.4
2.5
2.6
2.7
2.8
1
2
3
4
5
6
7
8
Serial Position within the Phrase

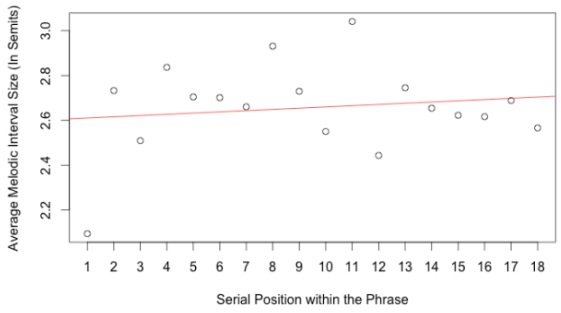

Average Melodic Interval Size (In Semits)
2.2
2.4
2.6
2.8
3.0
1
2
3
4
5
6
7
8
9
10
11
12
13
Serial Position within the Phrase

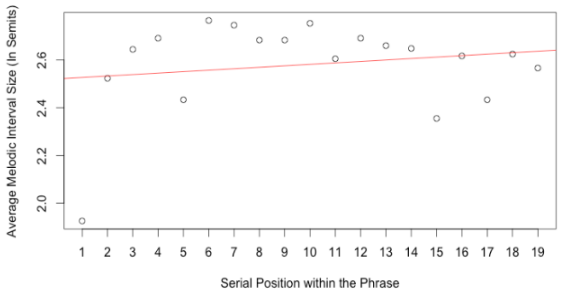

Average Melodic Interval Size (In Semits)
2.2
2.3
2.4
2.5
2.6
2.7
2.8
1
2
3
4
5
6
7
8
9
Serial Position within the Phrase

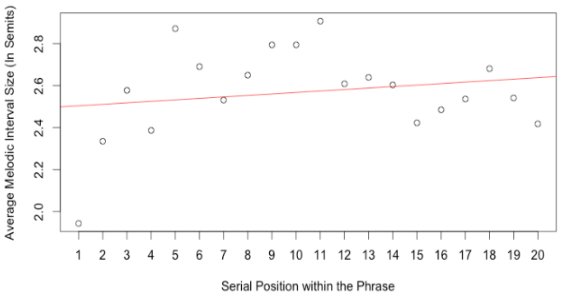

Average Melodic Interval Size (In Semits)
2.0
2.2
2.4
2.6
2.8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Serial Position within the Phrase


Average Melodic Interval Size (In Semits)
2.2
2.4
2.6
2.8
1
2
3
4
5
6
7
8
9
10
Serial Position within the Phrase


Average Melodic Interval Size (In Semits)
2.4
2.6
2.8
3.0
3.2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
Serial Position within the Phrase







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 (χ^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.

V. CONCLUSION

While the results of the first study are consistent with the experimental hypothesis, the results of the second study not. 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.

VI. 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.

It is possible that the initial interval size may 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 appear 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.

REFERENCES

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.

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.

Patel, A.D. (2008). *Music, Language, and the Brain.* Oxford: Oxford University Press.

Sachs, C. (1962). *The Wellsprings of Music.* The Hague: Martinus Nijhoff.

Shen, X.N. (1990) *The Prosody of Mandarin Chinese.* Berkeley: University of California Press.

Vassière, J. (1983). Language - independent prosodic features. In: A. Cutler and D.R.

Ladd, *Prosody: Models and Measurements.* Berlin: Springer, pp. 53-66.

Ladd, D.R. (2008). *Intonational Phonology,* 2nd edition. Cambridge: Cambridge University Press.

Schaffrath, H. (1995). *The Essen Folksong Collection.* D. Huron (ed.), Stanford, California: Center for Computer Assisted Research in the Humanities.

'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.