

Essen as a Corpus of Early Musical Experience

NIELS J. VEROSKY [1]
New York University Abu Dhabi

ABSTRACT: Statistics derived from the Essen Folksong Collection have widely been used as a proxy for general stylistic norms familiar to Western listeners. Since the specific facets of contemporary musical experience best modeled by a corpus of nineteenth-century European folksongs remain ambiguous, this study tests whether Essen-like music might be familiar to North American listeners through common children’s songs. Comparison with a corpus of 38 English-language children’s songs highly popular in North America finds that scale degrees from Essen and the children’s song corpus have near-perfect correlations in frequency profiles as well as high to very high correlations in tonal expectations and 4-grams. Profiles of scale degrees’ downbeat probabilities and average durations have moderate to high correlations for the diatonic but not the total chromatic. Overall, profiles of scale-degree behavior from the children’s song corpus match profiles from Essen more closely than do profiles from another corpus of music widely familiar to contemporary listeners (Billboard Hot 100 songs) and similarly closely as a corpus of nineteenth-century common-practice German vocal music (Schubert songs). For contemporary North American listeners, studies relying on Essen might plausibly be reinterpreted in terms of Essen acting as a corpus of early musical experience although the generalizability of Essen-derived statistics likely depends on the precise statistics being measured.

Submitted 2022 September 30; accepted 2023 May 26.

Published 2023 November 9; <https://doi.org/10.18061/emr.v17i2.9266>

KEYWORDS: *music, corpus, Essen, children’s songs, folk music*

THE Essen Folksong Collection (in the following, Essen) has been extensively used through several decades of research in music cognition, modeling, and information retrieval. Such a large body of research relying on Essen raises questions about which aspects of contemporary Western listeners’ musical experience can be captured by statistics taken from a corpus mostly comprising nineteenth-century folksongs.

Much work employing Essen involves comparing statistics derived from the corpus with data collected from human participants, making the relationship between Essen and the actual musical experience of participants in music cognition studies especially relevant. For example, models trained in whole or in part on Essen have been tested by Temperley (2008) on intervallic expectation data (Cuddy & Lunney, 1995), by Morgan, Fogel, Nair, and Patel (2019) on melodic cloze data (Fogel, Rosenberg, Lehman, Kuperberg, & Patel, 2015), and by Hansen, Kragness, Vuust, Trainor, and Pearce (2021) on melodic segmentation data. The interpretation of results from studies like these might vary depending on the degree to which statistics derived from Essen generalize beyond nineteenth-century European folksong.

Some work supports the common implicit or explicit assumption that characteristics of Essen reflect those of common-practice music or Western tonal music more broadly. Huron (2006) found that frequency profiles from Essen have a correlation of $r = .92$ with frequency profiles from Bach fugues while Arden (2003) reports a correlation of $r = .99$ with frequency profiles from common-practice art-song melodies compiled by Knopoff and Hutchinson (1983). Beyond common-practice music, Verosky (2021) found correlations of $r = .96$ for the total chromatic and $r = .85$ for diatonic scale degrees with major diatonic melodies from the de Clercq and Temperley (2011) rock corpus.

Nevertheless, research specifically focusing on Essen’s properties has been disproportionately limited in comparison with research drawing on Essen for other purposes (Brinkman, 2020), contributing to a situation where the interpretation of statistics taken from Essen remains ambiguous. As Brinkman (2021) writes, “It is rather unclear what exactly the Essen Collection is,” but Essen “seems to be a database



containing folksongs pulled from Germanic regions of Central Europe during the nineteenth-century.” Brinkman cautions that “scholars would do well to keep these fragile definitions in mind,” or risk a situation where “the repercussions derived from the number of assumptions made from working with the Collection become unmanageable.” Similar concerns that Essen’s use has outpaced our understanding of its generalizability date back at least to 2007, when Van Kranenburg et al. (2007) pointed out that none of the eight ISMIR papers employing Essen between 2001-2006 had any particular emphasis on folksong, mostly choosing the corpus for pragmatic reasons instead.

Here, I explore the possibility that statistics derived from Essen are neither simply a proxy for common-practice music, nor are they solely interpretable in terms of 19th-century European folksong. Rather, I hypothesize that, in the context of music cognition studies involving contemporary North American participants, highly popular children’s songs constitute the Essen-like music most familiar to listeners. Thus, when statistics derived from Essen are used to model statistical learning on the part of listeners, those statistics might most closely reflect statistical learning that takes place early in the process of music acquisition through popular children’s songs.

A primary motivation for this hypothesis is the observation that both melodies from Essen and highly popular North American children’s songs appear to be monophonic folksongs influenced by some general organizing principles of Western tonality (emphasis on dominant and tonic pitches, frequent use of a major diatonic scale, etc.). As Van Kranenburg and Janssen (2014) note, “an important property of folk melodies from Western oral tradition is the fact that such melodies were sung by ordinary people with little or no formal musical training.” Any resulting musical characteristics might be shared by Essen and contemporary North American children’s songs, or perhaps even exaggerated in melodies intended to be sung to and by children. In a global comparison, Sato et al. (2019) found that children’s and adults’ songs had similar intervallic characteristics, but that children’s songs were distinctive in featuring smaller melodic ranges and fewer scale degrees, suggesting sensorimotor constraints on melodic structure that may be especially pronounced in children’s songs.

Furthermore, some children’s songs popular in North America, such as *Twinkle, Twinkle, Little Star* (List, 1978), originate in traditional European songs. In contrast with melodies from Essen, though, a small number of highly popular English-language children’s songs appear to be widely recognized by children and adults in the United States and Canada (Harwood, 1987; Killian, 1996; Mang, 2007; McGuire, 2000; Ries, 1982). Consequently, when Essen-derived statistics are tested as predictors of data from music perception experiments involving contemporary North American listeners, the corpus statistics might be understood as best modeling music that stylistically resembles but is more familiar to listeners than Essen: highly popular children’s songs.

Of course, statistics from Essen are only representative of children’s songs to the extent that those statistics in fact mirror the same statistics taken from children’s songs. This analysis tests the hypothesis of Essen as a corpus of early musical experience by comparing five statistics relating to tonality in Essen with the same statistics from a collection of 38 highly popular North American children’s songs. To examine whether musical characteristics of Essen might plausibly be most familiar to contemporary North American listeners through well-known children’s songs, correlations between Essen and the children’s corpus are compared with correlations between Essen and two comparison corpora – a Schubert song corpus that might be expected to be stylistically proximate to Essen relative to other common-practice corpora, and a Billboard Hot 100 song corpus that might be expected to be representative of music widely known to contemporary listeners. If Essen-like music is familiar to contemporary North American listeners through common children’s songs, then Essen’s correspondence with the children’s song corpus should be 1) at least as strong as Essen’s correspondence with other potentially stylistically proximate corpora (such as the Schubert corpus) and 2) stronger than Essen’s correspondence with other corpora familiar to contemporary listeners (such as the Billboard Hot 100 corpus).

METHODS

Corpora

Four corpora were used: Essen, a corpus of English-language children’s songs popular in North America, a common-practice comparison corpus of Schubert vocal melodies, and a popular comparison corpus of Billboard Hot 100 songs.

To compile the children's song corpus, children's songs widely known in North America were identified from McGuire (2000), Mang (2005), and the U.S. Department of State's Sing Out Loud Children's Songs (n.d.) webpage.

McGuire aimed to identify folksongs known across multiple geographical regions of the United States. Searching for areas of overlap between surveys of songs commonly known to children by Ries (1982), Harwood (1987), and Killian (1996), McGuire compiled a list of songs that appeared across multiple surveys. Table 1 from McGuire (2000) lists 26 such songs, all of which were included in the children's song corpus for the present analysis.

Mang observed the singing behavior of children between the ages of 2 and 4 in Vancouver, Canada for 42 months. The sample was small ($n = 8$) but heterogeneous in terms of languages spoken (English = 8, Cantonese = 2, Mandarin = 2, Spanish = 1). Mang mentions five examples of common children's songs, such as *Eency*, *Weency Spider* and *Pop Goes the Weasel*, that seemed to be known to all participants in the study. All five songs were added to the children's song corpus.

The U.S. State Department's Sing Out Loud Children's Song page is a resource for English language learners, described as a collection of "popular children's songs in the USA." Examples include *Twinkle, Twinkle, Little Star* and *I'm a Little Teapot*. All 13 songs in the collection were used for the children's song corpus.

All songs identified were transcribed into MIDI format for the present analysis, yielding a corpus with 38 children's songs. For songs from the Sing Out Loud page, the recording of the song provided on that page was used as the version for transcription. Since McGuire (2000) and Mang (2005) did not provide canonical versions of children's songs, the version known to the author of the present study was used for transcription. Where overlap between the Sing Out Loud, McGuire, and Mang lists occurred, the Sing Out Loud recording was used. These lists yielded two separate songs with nearly identical melodies (*Twinkle, Twinkle, Little Star* and *The Alphabet Song*), both of which were included in the corpus. Only the first verse of each children's song (i.e., one repetition of the song's melody) was transcribed.

Although the resulting corpus size is modest ($n = 38$), there is a tradeoff between the size of the corpus and the familiarity of the songs in it – a much larger corpus would require the inclusion of songs less commonly known in North America. The goal was not to compile an exhaustive list of North American children's songs, but rather a collection representing some children's songs likely to be known to a large number of music cognition study participants in North America.

The Essen Folksong Collection was downloaded in **kern format (Schaffrath, 1995). All 6,249 mostly European folksongs listed in the README were used for the present analysis. The 2,231 Chinese folksongs added to Essen, which have frequently been omitted from studies intending to draw on Essen as a proxy for Western tonal music, were not used.

To provide a reference point for similarity between Essen and common-practice music, a corpus of 35 Schubert vocal melodies (KernScores, 2001) was also used. The Schubert corpus was chosen on the basis that, as a corpus of German vocal melodies composed in the nineteenth century, it might be expected to be stylistically proximate to Essen relative to other common-practice music (as found by Aarden (2003) with regards to correlations in scale-degree distributions between Essen and common-practice German vocal melodies) while remaining somewhat representative of common-practice music more widely given that Schubert is one of the most frequently encountered common-practice composers (London, 2013).

Additionally, the 215 melodic transcriptions of Billboard Hot 100 songs from 1959-1991 provided as part of the Coordinated Corpus of Popular Music (CoCoPops) were used as a reference corpus of popular music (Computational and Cognitive Musicology Lab, 2019; Gauvin et al., 2017). Although Billboard Hot 100 songs likely cannot be expected to be stylistically proximate to Essen to the same extent that nineteenth-century common-practice German vocal melodies from the Schubert corpus can, the Billboard corpus instead provides a reference point for similarity between Essen and popular music familiar to contemporary North American listeners. Thus, the Schubert corpus provides a baseline for comparison with a common-practice corpus that might be expected to be stylistically proximate to Essen while the Billboard corpus provides a baseline for comparison with a popular corpus that might be expected to be widely familiar to contemporary North American listeners.

Analysis

Statistics relating to scale degrees' corpus-wide tonal properties were extracted from all three corpora using the music21 Python library to read files in MIDI and **kern format (Cuthbert & Ariza, 2010). Table 1 shows

all statistics that were collected. Profiles of scale degrees' occurrence frequencies have been used for key finding (Krumhansl & Schmuckler, 1986; Albrecht & Shanahan, 2013 for a comparison of distributional key finding models) as well as predicting scale degrees' relative tonal stability (Krumhansl, 1990; Loui, Wessel, & Hudson Kam, 2010; Oram & Cuddy, 1995). They have also been incorporated into models of melodic expectation (Temperley, 2008).

Table 1. *Statistics collected for each corpus.*

Statistic	Description
Frequency of occurrence	Scale degree's corpus-wide frequency of occurrence
Scale degree expectations	Scale degree's temporal association with other scale degrees as measured by expectation networks
4-grams	Length-4 scale-degree sequence's corpus-wide frequency of occurrence
Downbeat probability	Probability that a scale degree appears on the downbeat
Average duration	Scale degree's corpus-wide average duration in quarter notes

Scale degrees' tonal associations as captured by expectation networks (Verosky, 2019) measure which other scale degrees tend to occur soon (but not necessarily immediately) after a given scale degree. This measure of temporal association between scale degrees has been shown to predict listeners' ratings of pitch similarity (Verosky, 2019) as well as listeners' expectations for how melodies will continue (Verosky & Morgan, 2021).

Four-grams represent frequencies for every possible sequence of four contiguous scale degrees (12^4 total). Whereas profiles of frequencies and scale degree expectations measure scale degrees' overall frequencies and temporal associations with other scale degrees respectively, 4-gram profiles measure which short sequences of scale degrees are typical of a given corpus. N -grams have previously been found to predict listeners' perceptions of uncertainty (Hansen & Pearce, 2014) and complexity (Sauvé & Pearce, 2019) in as well as electrophysiological responses to melodies (Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010). For the current analysis, $n = 4$ was chosen with the goal that the resulting sequences would be long enough to represent characteristic melodic patterns and short enough to avoid excessive n -gram sparseness.

Downbeat probability and average duration capture some aspects of scale degrees' rhythmic treatment. Prince and Schmuckler (2014) showed that scale degrees rated as more stable in the Krumhansl-Kessler tonal hierarchy (Krumhansl & Kessler, 1982) appear disproportionately often on downbeats, and Verosky (2021) found that a scale degree's tendency to occur on downbeats predicts its place in the tonal hierarchy above and beyond other variables like frequency. For the present analysis, a scale degree's downbeat probability was defined as its number of downbeat occurrences divided by its total number of occurrences (if the total number of occurrences was nonzero).

Scale-degree durations are incorporated into Krumhansl-Schmuckler and some derivative key finding algorithms, where they appear to improve the algorithms' performance (Madsen & Widmer, 2007). Duration also seems to be a salient cue of tonal stability for listeners (Lantz & Cuddy, 1998; Lantz & Cuddy, 2006; Smith & Schmuckler, 2004). Here, a scale degree's average duration is defined as the cumulative duration of all occurrences in quarter notes divided by the number of occurrences (if nonzero).

Profiles based on each of these five measures were built for the four corpora, resulting in corpus-specific vectors of length 12 (frequency, downbeat probability, and duration profiles), length 12^2 (scale degree expectation profiles), and length 12^4 (4-gram profiles). Since the children's song corpus contained exclusively songs in the major mode, only major-key excerpts from each corpus were used. Thus, each corpus yielded a set of five profiles representing all major-key excerpts in that corpus.

Correlations in corpus profiles were then computed between Essen and the children's song corpus as well as between Essen and the two comparison corpora.[2] Correlations were reported both for profiles of the total chromatic and for profiles of diatonic scale degrees since correlations for the total chromatic can disproportionately reflect the difference between in-scale and out-of-scale tones, leading to high values of r not necessarily representative of diatonic scale degrees (Verosky, 2021). Finally, permutation tests were performed (Good, 2005) to determine whether the correlation between Essen and the children's song corpus was significantly stronger than the correlations between Essen and the comparison corpora. Specifically, pieces were randomly shuffled between the children's song corpus and a given comparison corpus, then correlations with Essen were recomputed. This process was repeated 100,000 times for each comparison corpus, resulting in a one-tailed p -value for the hypothesis that the difference in correlations with Essen between the children's song corpus and comparison corpus is greater than zero (i.e., that the children's song corpus has a stronger correlation with Essen than the comparison corpus). Python code for the analysis can be accessed at <https://dx.doi.org/10.17605/OSF.IO/VAU5M>, as can the transcribed children's song corpus.

RESULTS

Essen's correlations with the children's song and comparison corpora are shown in Table 2.

Table 2. *Correlations between corpora.*

Statistic	Children's Song Corpus	Schubert Corpus	Billboard Corpus
Frequency of occurrence	r (chromatic) = 1.00	r (chromatic) = .99	r (chromatic) = .94
	r (diatonic) = .99	r (diatonic) = .95	r (diatonic) = .82
Scale degree expectations	r (chromatic) = .98	r (chromatic) = .96	r (chromatic) = .85
	r (diatonic) = .96	r (diatonic) = .90	r (diatonic) = .73
4-grams	r (chromatic) = .78	r (chromatic) = .68	r (chromatic) = .56
	r (diatonic) = .76	r (diatonic) = .66	r (diatonic) = .51
Downbeat probability	r (chromatic) = .22	r (chromatic) = .55	r (chromatic) = .02
	r (diatonic) = .76	r (diatonic) = .62	r (diatonic) = .41
Average duration	r (chromatic) = -.69	r (chromatic) = .22	r (chromatic) = -.60
	r (diatonic) = .60	r (diatonic) = .90	r (diatonic) = .43

For profiles of scale degrees' occurrence frequencies, temporal associations with other scale degrees, and four-gram frequencies, Essen's correlations with the children's song corpus range from high to near-perfect and are consistently as strong or stronger than corresponding correlations with Schubert songs and Billboard 100 songs.

Results for downbeat probability and average duration, both measures of scale degrees' rhythmic treatment, are more ambiguous. In particular, correlations with the children's song corpus are much lower for the total chromatic than the diatonic set. That pattern should be interpreted in light of the fact that non-diatonic scale degrees are extremely rare in the children's song corpus, with only 6 occurrences throughout the 38 children's songs.

Even considering only correlations for the diatonic, however, the duration profiles are anomalous in that Essen's correlation for children's songs ($r = .60$) is lower than for Schubert songs ($r = .90$), albeit still higher than for Billboard 100 songs ($r = .43$). Thus, scale degrees' duration profiles in Essen may not resemble corresponding profiles from children's song to the extent that other profiles considered here do.

Table 3 shows results of permutation tests for the hypothesis that the correlation between profiles from the children's song corpus and profiles from Essen is stronger than the correlation between profiles from each comparison corpus and profiles from Essen.

Table 3. *Permutation tests of whether the children's song corpus has stronger correlations with Essen than the Schubert and Billboard corpora do.*

Statistic	Children's Corpus r > Schubert Corpus r	Children's Corpus r > Billboard Corpus r
Frequency of occurrence	p (chromatic) = .17	p (chromatic) = .001**
	p (diatonic) = .20	p (diatonic) = .002**
Scale degree expectations	p (chromatic) = .06	p (chromatic) < .00001***
	p (diatonic) = .09	p (diatonic) < .00001***
4-grams	p (chromatic) = .006**	p (chromatic) < .00001***
	p (diatonic) = .02*	p (diatonic) < .00001***
Downbeat probability	p (chromatic) = .83	p (chromatic) = .36
	p (diatonic) = .21	p (diatonic) = .047*
Average duration	p (chromatic) = 1.0	p (chromatic) = .96
	p (diatonic) = .96	p (diatonic) = .25

Differences in correlations between the children's song corpus and the Billboard corpus are mostly highly significant while differences between the children's song corpus and the Schubert corpus are mostly nonsignificant. This pattern indicates that the children's song corpus corresponds more closely to Essen than Billboard Hot 100 songs do but similarly strongly as Schubert songs, consistent with the idea that North American children's songs are as closely related to Essen as other music that might be expected to be stylistically proximate to Essen (nineteenth-century common-practice German vocal melodies) and more closely related to Essen than other music familiar to contemporary North American listeners (Billboard Hot 100 hits).

Examining the magnitude of the correlations in Table 2, Essen's profiles match those from North American children's songs well for scale degrees' treatment in terms of frequency, expectations, 4-grams, and—for diatonic scale degrees—metric placement. For diatonic scale degrees' treatment in terms of duration, profiles from Essen still correlate reasonably well with those from children's songs, but not as well as with common-practice Schubert songs.

DISCUSSION

Overall, comparison of profiles between Essen and popular North American children's songs indicates that statistics derived from Essen may be representative of musical stimuli many North American listeners encounter in the initial stages of music acquisition. While children's songs constitute only a fraction of adult listeners' musical experience, children are already sensitive to key membership at 4-5 years of age (Corrigall & Trainor, 2010), diatonicity by 4-6 years of age (Trehub, Cohen, Thorpe, & Morrongiello, 1986), implied harmony by 7 years of age (Trainor & Trehub, 1994), and tonal hierarchy by 6-8 years of age (Cuddy & Badertscher, 1987). Therefore, songs widely sung to young children might play a disproportionately large role in developing familiarity with basic constructs of Western tonal music.

If highly popular children's songs represent an influential portion of listeners' early musical experience, then it appears that statistics derived from Essen might plausibly be reinterpreted in terms of early musical experience when used for comparison with experimental data from contemporary North American listeners. Given some variation in how closely profiles from Essen match those from the children's

song corpus, though, the generalizability of Essen-derived statistics may depend on exactly which statistics are being used.

The most straightforward case is when scale-degree frequency profiles from Essen are used. Since these profiles have a near-perfect correlation with corresponding profiles from children's songs for major keys, the results of any study relying on major-key frequency profiles from Essen would likely be the same if major-key frequency profiles from popular North American children's songs were used instead. Consequently, findings that pitch-class distributions from Essen provide sufficient information to reliably differentiate between major keys (Aarden, 2003; Huron, 2006; Temperley & Marvin, 2008) imply that distributions from highly popular children's songs provide equivalent information about key. The correlation between distributions from Essen and children's songs also offers preliminary but incomplete evidence for the generalizability of analyses that combine frequency profiles with other types of information extracted from Essen (e.g., Chan & Hsiao, 2016; Eerola, 2016; Rohrmeier, Rebuschat, & Cross, 2011; Temperley, 2008; Temperley, 2014).

Like frequency profiles, associations of which scale degrees tend to occur close together in time have very high correlations between Essen and children's songs. Verosky and Morgan (2021) have shown that temporal associations between scale degrees in Essen predict listeners' melodic expectations above and beyond *n*-grams or Temperley's (2008) intervallic model, a result which could now be reinterpreted as likely implying that temporal associations between scale degrees in children's songs supply equivalent information. Additionally, because temporal associations between scale degrees have a close relationship with transition probabilities between scale degrees (Verosky, 2019), the correlation in scale-degree expectations between Essen and children's songs supports the generalizability of research using scale-degree transition probabilities extracted from Essen (Chan & Hsiao, 2016; Huron, 2006; Temperley, 2014).

The high correlation in 4-gram profiles offers some indication that common four-note sequences of scale degrees in Essen typify such four-note sequences used in children's songs. *N*-grams from Essen have been used to model monophonic music (Bod, 2002; Pearce & Wiggins, 2004; Pearce & Wiggins, 2006) and predict listeners' melodic expectations (Hansen & Pearce, 2014; Hansen et al., 2021; Morgan et al., 2019; Pearce, Müllensiefen, & Wiggins, 2010), indicating that *n*-grams from children's songs could be used similarly. However, while the correlation for 4-grams between Essen and children's songs is high, it is not as high as the correlations for scale degree expectations and frequency profiles, perhaps either because 4-grams are more narrowly style-specific or because they naturally have more variation (as there are 12⁴ chromatic and 7⁴ diatonic 4-grams possible).

Moderate-to-high correlations in scale degrees' downbeat probabilities and duration profiles between Essen and children's songs suggest some resemblance in scale degrees' rhythmic treatment between the two corpora. However, this pattern may be limited to diatonic scale degrees, and Essen's duration profiles more closely correspond to those from Schubert songs than those from children's songs. Therefore, rhythmic statistics from Essen may not generalize to contemporary North American listeners' early musical experience to the same extent as statistics like scale degrees' frequencies, temporal associations, and 4-grams that abstract away from rhythm.

Since the corpus of popular children's songs contained only music in major keys, the present results speak most directly to the use of major-key statistics from Essen. When major-key statistics from Essen resemble major-key music familiar to children, it may or may not follow that the corresponding minor-key statistics from Essen resemble minor-key music familiar to children – depending, for example, on whether children's minor-key musical experience reflects similar sensorimotor constraints as or relies on extrapolating from their more extensive major-key musical experience. The lack of popular minor-key children's songs itself suggests interesting questions about differential acquisition of major-key and minor-key tonal knowledge, consistent with a major-mode bias in key perception (Temperley & Marvin, 2008).

Ultimately, it appears that several Essen-derived statistics mirror analogous statistics from songs widely known to contemporary North American listeners during music acquisition in childhood. Nonetheless, caution is needed when reinterpreting statistics from Essen in terms of early musical experience insofar as some Essen-derived statistics may generalize more than others.

ACKNOWLEDGEMENTS

This article was copyedited by Gabriele Cecchetti and layout edited by Diana Kayser.

NOTES

[1] Correspondence can be addressed to: Niels Verosky, niels.verosky@gmail.com.

[2] Corpus profiles are compared here using Pearson's correlation coefficient, as has been common in previous work comparing profiles of musical corpora. However, other methods of comparing corpus profiles, such as Kullback-Leibler divergence, cosine similarity, or mean absolute error, could be used.

REFERENCES

Aarden, B. J. (2003). *Dynamic melodic expectancy*. (Unpublished doctoral dissertation). Ohio State University.

Albrecht, J., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception*, 31, 59-67. <https://doi.org/10.1525/mp.2013.31.1.59>

Bod, R. (2002). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, 31, 27-36. <https://doi.org/10.1076/jnmr.31.1.27.8106>

Brinkman, A. (2020). *Exploring the structure of Germanic folksong*. (Unpublished doctoral dissertation). Ohio State University.

Brinkman, A. (2021, February). *Analyses on the history, form, and use of the Essen Folksong Collection* [Paper presentation]. South Central Society for Music Theory. https://www.scsmt.org/2021_conf_files/brinkman_2021_text.pdf

Chan, A.B., & Hsiao, J. H. (2016). Information distribution within musical segments. *Music Perception*, 34, 218-242. <https://doi.org/10.1525/mp.2016.34.2.218>

Computational and Cognitive Musicology Lab. (2019). CoCoPops-Billboard. Retrieved from <https://github.com/Computational-Cognitive-Musicology-Lab/CoCoPops-Billboard>

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, 28, 195-200. <https://doi.org/10.1525/mp.2010.28.2.195>

Cuddy, L. L., & Badertscher, B. (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience. *Perception & Psychophysics*, 41, 609-620. <https://doi.org/10.3758/BF03210493>

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics*, 57, 451-462. <https://doi.org/10.3758/BF03213071>

Cuthbert, M. S., & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. *Proceedings of the 11th International Symposium on Music Information Retrieval*, 637-42.

de Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, 30, 47-70. <https://doi.org/10.1017/S026114301000067X>

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review*, 11, 2-17. <https://doi.org/10.18061/emr.v11i1.4836>

Fogel, A. R., Rosenberg, J. C., Lehman, F. M., Kuperberg, G. R., & Patel, A. D. (2015). Studying musical and linguistic prediction in comparable ways: The melodic cloze probability method. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01718>

Gauvin, H. L., Condit-Schultz, N., & Arthur, C. (2017). Supplementing melody, lyrics, and acoustic information to the McGill Billboard Database. *Annual Conference of the International Alliance of Digital Humanities Organizations*.

Good, P. I. (2005). *Permutation, parametric and bootstrap tests of hypotheses* (3rd ed.). Springer.

Hansen, N. C., Kragness, H. E., Vuust, P., Trainor, L., & Pearce, M. T. (2021). Predictive uncertainty underlies auditory boundary perception. *Psychological Science*, *32*, 1416-1425. <https://doi.org/10.1177/0956797621997349>

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*. <https://doi.org/10.3389/fpsyg.2014.01052>

Harwood, E. E. (1987). *The memorized song repertoire of children in grades four and five in Champaign, Illinois*. (Publication No. 8721651) [Doctoral dissertation, University of Illinois at Urbana-Champaign]. ProQuest Dissertations Publishing. <https://www.proquest.com/openview/4845136f4a6c8bf1a19370fe5db770ee/>

Huron, D. (2006). *Sweet anticipation*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/6575.001.0001>

KernScores. (2001). Retrieved from <http://kern.ccarh.org/cgi-bin/ksbrowse?type=collection&l=/osu/monophony/schubert>

Killian, J. N. (1996). Definitions of “knowing”: Comparison of verbal report versus performance of children's songs. *Journal of Research in Music Education*, *44*, 215-228. <https://doi.org/10.2307/3345595>

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, *27*, 75–97. <https://doi.org/10.2307/843561>

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*, 334-368. <https://doi.org/10.1037/0033-295X.89.4.334>

Krumhansl, C. L., & Schmuckler, M. A. (1986, July). *Key-finding in music: An algorithm based on pattern matching to tonal hierarchies* [Paper presentation]. The 19th Annual Meeting of the Society of Mathematical Psychology, Cambridge, MA.

Lantz, M. E., & Cuddy, L. L. (1998). Total and relative duration as cues to surface structure in music. *Canadian Acoustics*, *26*, 56–57.

Lantz, M. E., & Cuddy, L. L. (2006). Can silence affect perception? Duration and frequency of occurrence in perceived pitch structure. *Canadian Acoustics*, *34*, 38–39.

List, G. (1978). The distribution of a melodic formula: Diffusion or polygenesis? *Yearbook of the International Folk Music Council*, *10*, 33-52. <https://doi.org/10.2307/767346>

London, J. (2013). Building a representative corpus of classical music. *Music Perception*, *31*, 68-90. <https://doi.org/10.1525/mp.2013.31.1.68>

Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, *27*, 377–388. <https://doi.org/10.1525/mp.2010.27.5.377>

Madsen, S., & Widmer, G. (2007). Key-finding with interval profiles. *Proceedings of the International Computer Music Conference*, 212–215.

- Mang, E. (2005). The referent of children's early songs. *Music Education Research*, 7, 3-20. <https://doi.org/10.1080/14613800500041796>
- McGuire, K. M. (2000). Common songs of the cultural heritage of the United States: A compilation of songs that most people "know" and "should know." *Journal of Research in Music Education*, 48, 310-322. <https://doi.org/10.2307/3345366>
- Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition*, 189, 23-34. <https://doi.org/10.1016/j.cognition.2018.12.015>
- Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, 39, 1367-1391. <https://doi.org/10.1068/p6507>
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage*, 50, 302-313. <https://doi.org/10.1016/j.neuroimage.2009.12.019>
- Pearce, M., & Wiggins, G. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, 33, 367-385. <https://doi.org/10.1080/0929821052000343840>
- Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, 23, 377-405. <https://doi.org/10.1525/mp.2006.23.5.377>
- Prince, J. B., & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis. *Music Perception*, 31, 254-270. <https://doi.org/10.1525/mp.2014.31.3.254>
- Ries, N. L. V. (1982). *An analysis of the characteristics of infant-child singing expressions*. (Publication No. 8223568) [Doctoral dissertation, Arizona State University]. ProQuest Dissertations Publishing. <https://www.proquest.com/openview/c7800c1565950fb7b675782044e2f4b7/>
- Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition: An International Journal*, 20, 214-222. <https://doi.org/10.1016/j.concog.2010.07.004>
- Sato, S., Six, J., Pfordresher, P., Fujii, S., & Savage, P. (2019). Automatic comparison of global children's and adult songs supports a sensorimotor hypothesis for the origin of musical scales. In I. Ali-MacLachlan & J. Hockman (Eds.), *Proceedings of the 9th International Workshop on Folk Music Analysis (FMA 2019)* (pp. 41-46). <https://doi.org/10.31234/osf.io/kt7py>
- Sauvé, S., & Pearce, M. T. (2019). Information-theoretic modeling of perceived musical complexity. *Music Perception*, 37, 165-178. <https://doi.org/10.1525/mp.2019.37.2.165>
- Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *The Essen folksong collection in **kern format*. Center for Computer Assisted Research in the Humanities. <http://kern.humdrum.org/cgi-bin/browse?l=/essen>
- Smith, N. A., & Schmuckler, M. A. (2004). The perception of tonal structure through the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 268-286. <https://doi.org/10.1037/0096-1523.30.2.268>
- Sing Out Loud Children's Songs. (n.d.). United States Department of State. <https://americanenglish.state.gov/resources/sing-out-loud-childrens-songs>

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, 56, 125–132. <https://doi.org/10.3758/BF03213891>

Trehub, S. E., Cohen, A. J., Thorpe, L. A., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 295–301. <https://doi.org/10.1037/0096-1523.12.3.295>

Temperley, D. (2008). A probabilistic model of melodic perception. *Cognitive Science*, 32, 418-444. <https://doi.org/10.1080/03640210701864089>

Temperley, D. (2014). Probabilistic models of melodic interval. *Music Perception*, 32, 85-99. <https://doi.org/10.1525/mp.2014.32.1.85>

Temperley, D., & Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception*, 25, 193–212. <https://doi.org/10.1525/mp.2008.25.3.193>

Van Kranenburg, P., Garbers, J., Volk, A., Wiering, F., Grijp, L., & Veltkamp, R. C. (2007). Towards integration of MIR and folk song research. *Proceedings of the 8th International Conference on Music Information Retrieval*.

Van Kranenburg, P., & Janssen, B. (2014). What to do with a digitized collection of Western folk song melodies? *Proceedings of the Fourth International Workshop on Folk Music Analysis*, 117-119.

Verosky, N. J. (2019). Corpus-based learning of tonal expectations with expectation networks. *Journal of New Music Research*, 48, 145-158. <https://doi.org/10.1080/09298215.2018.1542447>

Verosky, N. J. (2021). Interpreting the tonal hierarchy through corpus analysis. *Psychomusicology: Music, Mind, and Brain*. <https://doi.org/10.1037/pmu0000276>

Verosky, N. J., & Morgan, E. (2021). Pitches that wire together fire together: Scale degree associations across time predict melodic expectations. *Cognitive Science*, 45(10). <https://doi.org/10.1111/cogs.13037>