Title	Information distribution within musical segments
Author(s)	Chan, AB; Hsiao, JHW
Citation	Music Perception, 2016, v. 34 n. 2, p. 218-242
Issued Date	2016
URL	http://hdl.handle.net/10722/232935
Rights	Music Perception. Copyright © University of California Press.; Published as [provide complete bibliographic citation, as appears in the print version of your journal]. © [2016] by [the Regents of the University of California/Sponsoring Society or Association]. Copying and permissions notice: Authorization to copy this content beyond fair use (as specified in Sections 107 and 108 of the U. S. Copyright Law) for internal or personal use, or the internal or personal use of specific clients, is granted by [the Regents of the University of California/on behalf of the Sponsoring Society] for libraries and other users, provided that they are registered with and pay the specified fee via Rightslink® or directly with the Copyright Clearance Center.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Information Distribution Within Musical Segments

Antoni B. Chan
City University of Hong Kong, Kowloon Tong, Hong Kong

Janet H. Hsiao
University of Hong Kong, Pok Fu Lam, Hong Kong

In research on word recognition, it has been shown that word beginnings have higher information content for word identification than word endings; this asymmetric information distribution within words has been argued to be due to the communicative pressure to allow words in speech to be recognized as early as possible. Through entropy analysis using two representative datasets from Wikifonia and the Essen folksong corpus, we show that musical segments also have higher information content (i.e., higher entropy) in segment beginnings than endings. Nevertheless, this asymmetry was not as dramatic as that found within words, and the highest information content was observed in the middle of the segments (i.e., an inverted U pattern). This effect may be because the first and last notes of a musical segment tend to be tonally stable, with more flexibility in the first note for providing the initial context. The asymmetric information distribution within words has been shown to be an important factor accounting for various asymmetric effects in word reading, such as the left-biased preferred viewing location and optimal viewing position effects. Similarly, the asymmetric information distribution within musical segments is a potential factor that can modulate music reading behavior and should not be overlooked.

Received: August 22, 2013, accepted April 6, 2016.

Key words: Entropy analysis, musical segments, music reading, information distribution, optimal viewing position

In speech recognition, it has been shown that word beginnings usually convey more information than word endings in terms of entropy from information theory (Shannon, 1948). In other words, there is greater uncertainty/variability at word beginnings, and thus it is easier to differentiate words using word beginnings than word endings. For example, Yannakoudakis and Hutton (1992) analyzed words in a large lexicon with 11,031 different words obtained from six very different texts and transcribed them into phonetic codes (Elovitz, Johnson, McHugh, & Shore, 1976; Yannakoudakis & Hutton, 1987); they found that in general, beginning positions in the words had higher entropy (i.e., higher information content) than ending positions, and that short words generally had higher entropy than long words (cf. Bourne & Ford, 1961). Shillcock, Hicks, Cairns, Charter, and Levy (1996) used a phonological transcription of the London-Lund Corpus of spoken English, a corpus of orthographically transcribed conversational English speech that contains more than 450,000 word tokens (Svartvik & Quirk, 1980), and showed that in general beginning segments of spoken words have higher information content than ending segments. This asymmetric information distribution is also reflected in written English words. For example, Shillcock, Ellison, and Monaghan (2000) calculated the entropy distribution across different letter positions with left-justified English words taken from the CELEX lexical database (Baayen, Pipenbrock, & Gulikers, 1995; in total 34,154 words containing derived but not inflected words); they showed that the entropy gradually decreased from beginning positions to ending positions. Consistent with this observation, in English there are more suffixes than prefixes (Carstairs-McCarthy, 2002; words with suffixes typically have more information in the word beginning; vice versa for those with prefixes). It has been argued that this asymmetric information distribution in English words is due to a communicative pressure to maximize the amount of information in word beginnings (or more specifically, to increase the variability of word beginnings) so that spoken words can be recognized efficiently before the end of the pronunciation, allowing time for other processes such as syntax processing (e.g., Brysbaert & Nazir, 2005; Shillcock, et al., 2000).

The asymmetric information distribution in English words also influences how people read written words. In reading isolated English words with a single fixation, it has been shown that people have the best word recognition performance when their fixation is initially directed to the left of the word center, closer to the word beginning than the word end (the optimal viewing position,

Music Perception, volume 34, issue 2, pp. 218–242, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2016.34.2.218

OVP; O'Regan, 1990; O'Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984). This asymmetric pattern has also been observed in reading continuous texts: readers most often fixate on word beginnings (the preferred viewing location, PVL; Rayner, 1979; see also Ducrot & Pynte, 2002; note that in English words the PVL is slightly more to the left than the OVP; Legge, Klitz, & Tian, 1997). The leftward biased OVP and PVL phenomena in English word reading have been proposed to be related to the asymmetric information distribution within words, in addition to the possible influence from left hemisphere lateralization in language processing and reading direction (e.g., Brysbaert & Nozir, 2005; Brysbaert, Vitu, & Schroyens, 1996; Legge et al., 1997).

Similar to speech, music is a medium of communication. Although an exact analogy cannot be drawn between the structures of speech and music, musical notes may be considered as analogous to phonemes in speech, while musical segments (e.g., a self-contained music fragment, a motif) and musical phrases (e.g., an 8-bar melody) are analogous to words and sentences. It remains unclear whether an asymmetric information distribution can be found within musical segments. In contrast to English words, musical segments do not follow strict morphological/orthographical rules, and do not have clearly defined segment boundaries and meanings. Music is frequently considered art and a form of creativity, and thus the structures of musical segments in songs may vary significantly across different songwriters (see, e.g., Knopoff & Hutchinson, 1983; Youngblood, 1958). Nevertheless, some consistent patterns of information structure may exist in musical segments. For example, melodies in Western music typically end with a tone that is stable (e.g., the perfect cadence) and thus is more predictable (Aarden, 2003), suggesting that there may be more information in musical segment beginnings than endings. Consistent with this speculation, Wong and Hsiao (2012) observed that in reading musical segments with a single fixation, musicians had better performance when their fixation was directed to musical segment beginnings than to endings (i.e., an asymmetric OVP pattern), suggesting that musical segment beginnings may have more information content for segment identification than endings. An examination of information distribution within musical segments not only will promote our understanding of how music is produced, but also the way we perceive, process, and perform music. For example, if musical segments have an asymmetric information distribution, musicians may consequently look at the side of a musical segment with higher information content more often when reading music scores. In contrast, if musical segments have a symmetric information distribution, the asymmetric OVP pattern observed in music reading (Wong & Hsiao, 2012) is unlikely to be due to the information distribution within musical segments. Thus, this examination will help us tease apart confounding factors that may influence eye fixation patterns in reading (Brysbaert & Nazir, 2005). In addition, the knowledge of information distribution within musical segments has important implications for studies of music perception, music acquisition, and human communication.

Another line of research focuses on discovering the regularities underlying the transitions of musical notes (e.g., Abdallah & Plumbley, 2009; Conklin & Witten, 1995; Pearce & Wiggins, 2006; Pearce, Ruiz, Kapasi, Wiggins, Bhattacharya, 2010), which promotes the understanding of melodic structures and the influence of statistical learning of these structures on music acquisition and expectation (e.g., Krumhansl & Kessler, 1982; Pearce & Wiggins, 2006; Rohrmeier & Rebuschat, 2012; Witten, Manzara, & Conklin, 1994). Most computational models of note transitions are based on *n-gram* models, where a conditional probability distribution predicts the n^{th} note given the $n-1$ preceding notes. A note can be represented by its pitch only (Abdallah and Plumbley, 2009), or in conjunction with other musical features, e.g., rhythm, onset, interval, etc. (Conklin & Witten, 1995; Pearce & Wiggins, 2006). In Pearce and Wiggins' IDyOM model, the prediction of a note is based on both a *long-term* model, which is estimated from a training corpus and reflects a person's prior knowledge of musical patterns, and a *short-term* model, which is estimated from the previous notes in the current melody and reflects the person's adaptation to the current melodic context (Conklin & Witten, 1995; Pearce & Wiggins, 2006). N-gram models have been used to explain human data of music perception. For example, Abdallah and Plumbley (2009) use the predictive information rate (conditional mutual information) as a measure of "surprise," while Witten et al. (1994), Pearce and Wiggins (2006), and Pearce, Ruiz, Kapasi, Wiggins, and Bhattacharya (2010) found similarities between the entropies of the conditional distributions of predicted notes and human note expectancy (measured in entropy). The n-gram models in the previous studies are typically based on short sequences of notes (e.g., 2 or 3), not whole musical segments, and on measuring the entropy of the *conditional* distribution of the predicted note given the previous notes. Hence, none of these previous studies examined the overall information distribution within musical segments, as measured by the entropy of each position in the sequence (similar to

words). The overall information distribution within musical segments, and its consequences for how people perceive music, remain unclear.

In the research on music perception, it has been proposed that listeners' melodic expectations are influenced by two distinct cognitive systems: one is an innate and universal bottom-up perception system governed by Gestalt-like principles, whereas the other is a top-down system that is influenced by experience with music in different styles (i.e., the implication-realization theory, or IR theory; Narmour, 1990, 1992). While the nature of the innate mechanism remains controversial (see, e.g., Elman et al., 1996; Pearce & Wiggins, 2006), it has been consistently reported that experience with music structures modulates music perception. For example, Trainor and Trehub (1992) showed that adult listeners of Western tonal melodies performed better in detecting a change in one note when it was outside the key than when it was within the key; in contrast, infants (who did not have as much experience with Western tonal melodies) performed equally well in the two cases. In another study, Trainor and Trehub (1993) showed that the advantage in discriminating a melody change in the context of related keys over unrelated keys was observed in both prototypical and non-prototypical Western melodies in infants; in contrast, this advantage was observed only in prototypical but not in non-prototypical Western melodies in adults. These studies suggest a modulation effect of experience with Western tonal melodies on music perception (see also Trainor & Trehub, 1994). Thus, the information of statistical properties of music may be important for the understanding of effects of experience in music perception.

In the current study, we aim to investigate statistical properties of music through examining the information distribution within musical segments. More specifically, we analyze two large databases of over 13,000 songs (obtained from the Essen folksong dataset and Wikifonia, www.wikifonia.org) and examine the information distribution within musical segments of Western tonal music. Here by "musical segment" we mean the lowest level of the grouping structure of music (Lerdahl & Jackendoff, 1983). We consider musical segments predicted by four automatic methods, which are based on various principles of music perception, as well as musical segments annotated by humans. We then calculate the entropy and conditional entropy at different note positions separately for musical segments of different lengths (cf. Shillcock et al., 2000; Yannakoudakis & Hutton, 1987), and examine whether the information distribution within musical segments has asymmetric patterns similar to those observed in English words in speech. It should be noted that the identification of the lowest-level groupings tends to be ambiguous and subjective, as it is sometimes not clear where a group starts or ends. The segmentation methods used here may not always identify the lowest-level grouping, or even the same level of grouping. Nonetheless, asymmetric patterns in the information distribution of musical segments may appear in multiple levels of grouping, and thus it is constructive to consider several segmentation methods.

Method

SONG DATASETS

The current study is based on two song datasets, the Essen folksong corpus and the Wikifonia corpus. To facilitate a meaningful analysis, only songs written in major keys (according to the metadata in the datasets) were selected. The Essen folksong corpus (Schaffrath, 1995) consists of 7,704 transcribed folksongs, and the Wikifonia dataset consists of 5,843 transcribed songs downloaded from Wikifonia (www.wikifonia.org), which is a community-run database of "music lead sheets." Each song contains the monophonic melody and metadata, such as musical key and time signature. Each song in the Essen corpus contains human annotations of musical segments, whereas the Wikifonia corpus does not. The distributions of songs over different regions for Essen and different genres for Wikifonia are listed in Table 1. In the Essen dataset, about 60% of the songs are from German folksongs, followed by 29% from China. In the Wikifonia dataset, about one third of the songs were in the "jazz" or "pop" categories, and most of the songs are in popular genres. In the Essen dataset, the median song length was 47 notes and the average length was 53 notes with standard deviation of 30. The lengths of songs ranged from 8 to 502, with 95% of songs between 21 and 126 notes. In the Wikifonia dataset, the median song length was 153 notes and the average length was 174 notes with standard deviation of 102. The lengths of songs ranged from 8 to 1050, with 95% of songs between 45 and 421 notes. A song can be written in any musical key, e.g., to fit the target instruments. In order to facilitate a meaningful analysis of the notes relative to the key (the root note, tonic), each song was transposed into the common key of C major using the key information provided in each song file. Songs in minor keys were excluded in the analysis.

MUSIC SEGMENTATION

Each song melody was automatically segmented into a set of musical segments, consisting of short contiguous

TABLE 1. *Distribution of Songs in the Essen and Wikifonia Datasets According to Genre Labels*

Essen				Wikifonia	
America - Mexico	4	Europa - Lothringen	42	blues	91
America - misc	2	Europa - Luxemburg	8	broadway	437
America - USA	7	Europa - Hungary	34	children	40
Asia - China	2238	Europa - misc.	24	classic	207
Asia - misc.	3	Europa - Netherlands	51	folk	302
Europa - Czech	34	Europa - Austria	103	holiday	198
Europa - Denmark	3	Europa - Poland	15	jazz	1171
Europa - Germany	4755	Europa - Romania	21	movies	435
Europa - Alsace	87	Europa - Russia	33	none	1177
Europa - England	3	Europa - Switzerland	85	pop	948
Europa - France	9	Europa - Sweden	2	rock	186
Europa - Italy	7	Europa - Tirol	14	television	29
Europa - Yugoslavia	108	Europa - Ukraine	12	traditional	370
		TOTAL	**7704**	**TOTAL**	**5843**

groups of musical notes, i.e., the lowest-levels of the grouping structure (Lerdahl & Jackendoff, 1983). The musical segments are analogous to words in speech, and the notes analogous to phonemes in speech. In the literature, there have been studies on cognitive modeling of word segmentation using probabilistic approaches (e.g., Brent, 1999a, 1999b; Cohen, Adams, & Heeringa, 2007; Saffran, Newport, Aslin, Tunick, & Barrueco, 1996). Since the perception of music is subjective, there have been many algorithms proposed to segment music into groups of notes, which are based on different underlying principles. Here we considered four automatic approaches, of varying complexity, to segment each song. For Essen, we also use the human annotations of note groupings.

Temporal proximity (TP). We define a musical segment as a set of notes in close temporal proximity. The assumption is that longer time intervals between notes indicate pauses or focal points in the melody, which in turn indicate the end of a musical segment, and a beginning of a new one. Specifically, a note with an interonset interval (IOI)[1] longer than a threshold T forms the beginning of a musical segment. We define the threshold T as the main beat (tactus) induced by the time signature (meter) of the song. Most of the time signatures use the quarter note as the main beat. The exception is with compound meters (e.g., 9/8), where the dotted-quarter note is assumed to be the main beat, and hence the threshold is three beats.

[1] The interonset interval (IOI) of a note is defined as the time interval between the onset of the note and that of the previous note. The IOI includes the duration of the previous note and the rest between the previous note and the note.

The TP method is conceptually similar to the 2nd Grouping Preference Rule (GPR2b) of the Generative Theory of Tonal Music (GTTM, Lerdahl & Jackendoff, 1983). The main difference is that TP uses an absolute threshold of the IOI for determining the segment boundary, while GPR2b uses a threshold relative to the IOIs of the neighboring notes.

Local boundary detection model (LBDM). The LBDM by Cambouropoulos (1997) is based on detecting boundaries between musical segments using the relative change in three note properties: IOI, pitch interval, and rest time (time between offset of a note and onset of a new note). The probability of a boundary at a particular note is the weighted sum of the relative changes with its neighbors. We used the implementation of LBDM from the MIDI toolbox software package (Eerola & Toiviainen, 2004), and set the probability threshold for a boundary to 0.4, as suggested by experiments by de Nooijer, Wiering, Volk, & Tabachneck-Schijf (2008).

Grouper (GRP). The Grouper model was introduced by Tempereley (2001) and calculates a grouping of the melody using a set of Phrase Structure Preference Rules (PSPRs), which are based on temporal proximity, preferred phrase length, and consistency in relation to the meter. The note features used by Grouper consist of onset time, off time, chromatic pitch, and level in the metrical hierarchy. We used the Melisma Music Analyzer (Sleator & Temperley, 2003) to calculate the metrical hierarchy and Grouper segmentation, using the default parameters.

Information dynamics of music (IDyOM). The IDyOM model was proposed by Pearce, Müllensiefen, and Wiggins (2010), and is based on the principle that group

TABLE 2. *Comparison of Segmentation Methods on the Essen Folksong Corpus*

	Reference segments														
	GRP			TP			LBDM			IDyOM			H		
	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R
GRP	–	–	–	.63	.64	**.81**	.69	.64	**.86**	.54	.41	**.93**	.65	.64	.69
TP	.63	**.81**	.64	–	–	–	.72	.80	**.83**	.64	.59	**.91**	.58	.74	.60
LBDM	.69	**.86**	.64	.72	**.83**	.80	–	–	–	.66	.58	**.93**	.65	.78	.62
IDyOM	.54	**.93**	.41	.64	**.91**	.59	.66	**.93**	.58	–	–	–	.54	**.91**	.42
H	.65	.69	.64	.58	.60	.74	.65	.62	.78	.54	.42	**.91**	–	–	–

Note: In each column, segments from different segmentation methods are used as the reference, to which the F-measure, precision (P), and recall (R) of the other methods are calculated. Bold values indicate high levels of precision or recall (>.80).

boundaries are perceived before events that are unexpected given the context of the melody. Specifically, the model estimates the conditional probability distribution of a note given all previous notes, $p(x_i|x_{i-1}, \ldots, x_1)$, and calculates its *self-information* (or *surprisal*), $h(x_i|x_{i-1}, \ldots, x_1) = -\log_2 p(x_i|x_{i-1}, \ldots, x_1)$, which is a measure of unexpectedness or surprise of the note. Group boundaries are indicated by high values of self-information, relative to its linearly decaying weighted average. We used the implementation provided by the IDyOM project (Pearce, 2014) to estimate the conditional probability distributions[2] of a note's features (chromatic pitch, IOI, offset-onset interval) on each dataset. On Essen, we use 50^{th} order model (i.e., 50 notes are used as sequential context), while a 20^{th} order model was used for Wikifonia. In the next section our analysis of information content in musical segments is based on the entropy of scale degrees in segments. As entropy is the expected value of self-information, IDyOM segments may naturally contain high entropy (information content) in the beginning of their segments. Note that IDyOM is based on different note features (chromatic pitch, IOI, and offset-onset interval vs. scale degrees) and model order (50^{th} or 20^{th} vs. 0^{th} or 1^{st}) from our entropy analysis, and hence this effect will be tempered somewhat.

Human annotations (H). The Essen corpus provides human annotation of musical phrases in each song. The phrases are non-overlapping and contiguous, and thus form a grouping structure of the song. Note that using these annotations does not resolve the subjectivity or ambiguity of groupings, since it only represents one person's intuition about a song.

[2] Specifically, we learn the IDyOM "long-term model" on the original (non-transposed) songs. This gave slightly better results than using the transposed songs.

We applied the above segmentation methods to the two musical datasets. We first quantified the agreement (or disagreement) between the segmentation methods. The segments of one method are used as the "reference segmentation," to which the other segmentation methods are compared. Specifically, the boundary notes predicted by a segmentation method are compared with the boundary notes of the reference segmentation via precision (P), recall (R), and F-measure. Precision is the percentage of boundary note predictions that match a reference boundary note, while recall is the percentage of reference boundary notes that were predicted correctly. F-measure is the harmonic mean of precision and recall. Table 2 shows the P, R, and F values when each segmentation method is the reference on the Essen corpus.

To determine the relationship among the 4 automatic methods, consider the following two observations. First, when method A has low recall and high precision against reference method B, it indicates that A's boundary notes are aligned with B's boundary notes (high precision), but method A does not predict some of B's boundary notes (low recall). In other words, A's boundary notes are a subset of B's. Second, when method A has high recall and low precision against reference method B, it indicates that method A predicted all boundary notes of B (high recall) but with some extra predictions not found in B (low precision), and therefore the boundary notes of B are a subset of A's. Using these two observations, an examination of the precision and recall values in Table 2 suggests that the predicted boundary notes of the automatic segmentation methods form a nested set (up to some noise). TP, LBDM, and IDyOM have high precision (> 0.8) and relatively lower recall (< 0.7) when GRP is the reference method, which suggests that the majority of boundary notes of TP, LBDM, and IDyOM are a subset of GRP's boundary notes. TP has high recall (> 0.8) when GRP is the

TABLE 3. *Statistics of the Musical Segments Extracted Using the Segmentation Methods*

Essen	GRP	TP	LBDM	IDyOM	H
Total number	45,841	37,288	32,257	16,115	43,049
Maximum length	16	30	31	37	22
Average length	8.93	8.12	10.51	15.89	9.33
Standard deviation	2.26	6.23	6.75	8.51	3.31
Median length	9	6	9	14	9
Wikifonia	**GRP**	**TP**	**LBDM**	**IDyOM**	
Total number	109,165	157,318	45,920	34,802	
Maximum length	19	28	48	53	
Average length	9.31	5.65	14.44	19.33	
Standard deviation	2.77	4.16	10.48	12.33	
Median length	9	4	11	16	

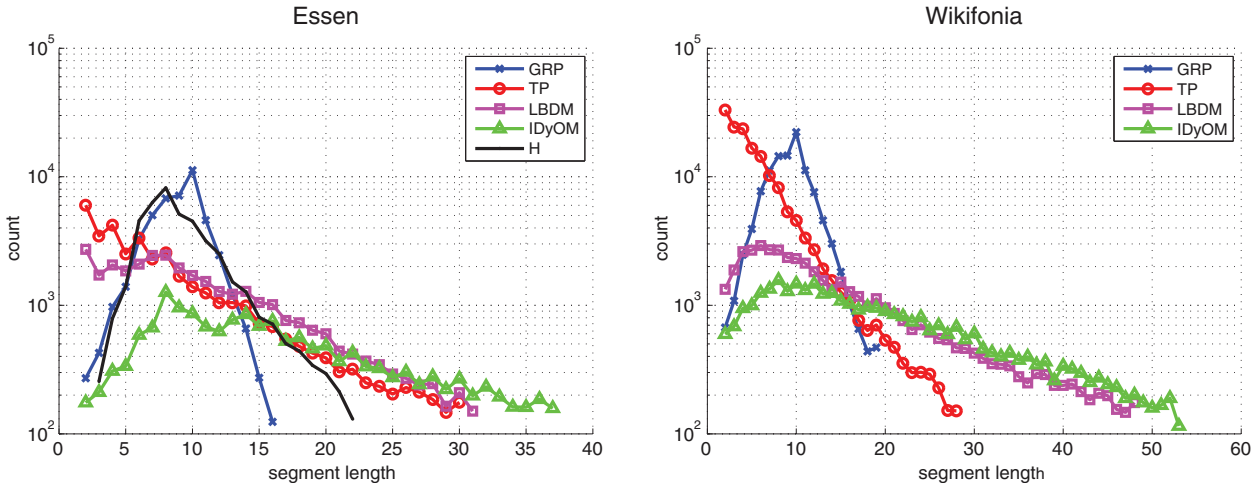

FIGURE 1. Distribution of musical segments of different lengths using the segmentation methods and human phrase annotations.

reference, indicating that most of TP's boundary notes are a subset of GRP. Likewise, the boundary notes of LBDM are mostly a subset of TP and GRP (recall both over 0.8). Finally, most of the boundary notes of IDyOM are subsets of all three methods (recall all over 0.9). The nested set of boundary notes suggests that each segmentation method identified a different level of the grouping structure, with GRP at the lowest-level (shorter segments), followed by TP and LBDM at the next two higher-levels, and finally IDyOM at the highest level (longest segments).

Compared to the human annotations, GRP has the highest recall and lowest precision among the segmentation methods, which suggests that GRP can identify more of the human annotated boundary notes but also predicts more boundary notes that do not agree with the human annotation. In contrast, IDyOM has the highest precision and lowest recall, which suggests that IDyOM predicts boundary notes more conservatively, but any predictions tend to agree with the human annotation. LBDM and TP are in between, but more similar to IDyOM, in that the precision is higher than recall.

For each method, the musical segments were grouped according to their lengths. Segment length groups with less than 144 samples were discarded, since there would not be enough samples to reliably estimate the note probabilities for those lengths. Table 3 presents the statistics of the extracted musical segments on the two datasets. Overall, TP and GRP tend to parse the melody into large sets of short segments (average lengths between 5 and 9). In contrast, LBDM and IDyOM segment the melodies into smaller sets of long segments (average lengths between 10 and 19). Figure 1 plots the total numbers of musical segments of different lengths

found using each segmentation method (see online PDF for color versions of all figures). The distributions are heavily concentrated on short sequences. A similar phenomenon was also observed in language; for example, according to an English word database developed by Brysbaert and New (2009), among the most frequent 25,000 (written) English words in the database, the lengths of the words range from 1 to 18, with the average length 7.17 and the median length 7.

The analysis of English words in Yannakoudakis and Hutton (1992) considered *unique* words extracted from a variety of sources (i.e., duplicate words are removed from the corpus). In language, there are specific rules about what letter combinations can appear together, which are reflected in the spelling of words. Music also has similar rules about what notes sound better together (more pleasing, less dissonant) in a musical segment. However, these are not hard rules, and hence any combination of notes could be played in a segment. Nonetheless, "good" note combinations will appear more frequently in music, and hence these musical rules can be inferred by considering *all* musical segments present in the dataset. That is, in this study, we do not restrict our analysis by removing duplicate musical segments. Rather, we feel it is more representative to look at all the musical segments in the dataset in order to infer its information distribution. Estimation from all segments also fits well with ideas from implicit learning of music, where it is theorized that a person acquires statistical models of note patterns through exposure to music throughout their lifetime (Rohrmeier & Rebuschat, 2012).

ENTROPY AND CONDITIONAL ENTROPY

Entropy is a measure of information content (Shannon, 1948): higher entropy indicates more information content, or in other words, more uncertainty/unpredictability. It has been shown to be able to capture several behavioral phenomena related to how humans process sequences of sensory input, such as language and music (e.g., Knopoff & Hutchinson, 1983; Reichle, Rayner, & Pollatsek, 2003; Shillcock et al., 2000). For example, in music perception, entropy and its related measures have been used as reflecting perceivable musical style (e.g., Knopoff & Hutchinson, 1983; Youngblood, 1958) and for modelling music listeners' internal representation of music structures and musical expectations (e.g., Abdallah & Plumbley, 1999; Pearce et al., 2010; Pearce & Wiggins, 2006). Thus in the current study we used entropy as the measure to uncover the information distribution of musical segments in the song datasets.

It should be noted that entropy is a property of a statistical distribution that is assumed to model the data source. In their analysis of English words, Yannakoudakis and Hutton (1992) calculated the entropy assuming a zeroth-order (unigram) model to represent the frequency of phonemes in each position of the words (i.e., the context around the position is not considered). In research on musical expectation, higher-order models are typically assumed (i.e., the context of the previous notes is included) since the aim is to measure the expectedness of a note while listening to a melody (e.g., Conklin & Witten, 1995; Manzara, Witten, & James, 1992; Pearce & Wiggins, 2006; Witten et al., 1994). In our analysis, we will consider both the zeroth-order (unigram) model, in order to parallel the linguistics study, as well as a first-order (bigram) model, following research on musical expectation. Due to lack of data, it was not possible to reliably estimate models with orders larger than 1.

For each set of musical segments of a given length, we calculated the entropy of notes at each position in the segment. We represent each note with its scale degree, i.e., its relationship with the tonic note. We define $\chi=\{1, \#1, 2, b3, 3, 4, \#4, 5, b6, 6, b7, 7\}$ as the set of 12 scale degrees, where we use integers 1 through 7 for the major scale degrees, with 1 as the tonic. For the zeroth-order model, we denote the probability of each of the 12 scale degrees in the i-th position ($i = 1, \ldots, L$) as $p(x_i^L)$, where $x_i^L \in \chi$ is the random variable of the scale degree at the i-th position in a length L segment. The probabilities are estimated using the relative frequency of occurrence in all length-L segments in the dataset. The entropy at each position $i = 1, \ldots, L$ is then calculated as

$$H(x_i^L) = -\sum_{j \in \chi} p(x_i^L = j) \log_2 p(x_i^L = j). \qquad (1)$$

The entropy is a measure of the randomness in a probability distribution, in this case the distribution of scale degrees at a particular position. A value of $H_{\min}=0$ indicates no randomness, e.g., a single scale degree is always played, whereas the maximum value of $H_{max} = \log_2 12 \approx 3.58$ indicates a uniform distribution, i.e., all scale degrees are equally likely. Since the maximum value of entropy is bounded, we define the normalized entropy as

$$\hat{H}(x_i^L) = \frac{H(x_i^L)}{H_{max}} \qquad (2)$$

which takes values from 0 to 1.

For the first-order model, we denote the conditional probability of the i-th note in a length L segment as

$p(x_i^L|x_{i-1}^L)$, where x_{i-1}^L is the previous note in the segment. The *specific conditional entropy* is defined as the entropy of the conditional distribution when the previous note is known and takes a *specific* value $x_{i-1}^L = k$,

$$H(x_i^L|x_{i-1}^L = k) = \\ -\sum_{j\in\chi} p(x_i^L = j|x_{i-1}^L = k)\log_2 p(x_i^L = j|x_{i-1}^L = k). \quad (3)$$

The *conditional entropy* is then defined as the specific conditional entropy averaged over all possible values of the previous note (Cover & Thomas, 1991),

$$H(x_i^L|x_{i-1}^L) = \sum_{k\in X} p(x_{i-1}^L = k)H(x_i^L|x_{i-1}^L = k), \quad (4)$$

where $p(x_{i-1}^L)$ is the probability distribution of the previous note at position $i-1$. The conditional entropy in Equation 4 is a measure of the uncertainty (information content) in the i-th note when the previous note $(i-1)$ is known. Similar to normalized entropy, we define the normalized conditional entropy as

$$\hat{H}(x_i^L|x_{i-1}^L) = \frac{H(x_i^L|x_{i-1}^L)}{H_{max}} \quad (5)$$

which ranges from 0 to 1. If the normalized conditional entropy is 0, then the i-th note is completely determined by the $(i-1)$-th note.

To compare with the information distribution of English words, we conducted similar analyses with the data from Yannakoudakis and Hutton (1992).[3] According to Rothschild (1986), the distribution of written English word lengths (in terms of number of letters) can be fitted with a shifted Poisson distribution with the mean 6.94 and the variance 5.80 letters (see also Bagnold, 1983). Although in Yannakoudakis and Hutton's (1992) data, word length information was based on number of phonemes instead of letters, we used Rothschild's (1986) data of written words as an estimate of a representative sample of English words and analyzed the data of words with lengths ranging from 2 to 12 in Yannakoudakis and Hutton's (1992) data (i.e., the mean word length minus/plus two standard deviations according to Rothschild, 1986).

In the above analysis, we used scale degrees to represent each note in order to align with the prior analyses of English letters/phonemes. On the other hand, in music, relative pitch, i.e., the pitch interval between two consecutive notes, is also important for mental encoding and recognition of melodies (e.g., Cuddy & Cohen, 1976; Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Peretz & Babaï, 1992). Hence, in a second analysis, we represent each note in a musical segment by the pitch interval, in semitones (half steps), between the note and its preceding note. The first note in the musical segment is ignored since it has no preceding note in the segment. Intervals that are an octave or greater (less than −11 or greater than +11) are mapped back to within one octave, while keeping the same decreasing/increasing direction. We define the set of 23 pitch intervals as $\chi = \{-11, \ldots, -1, 0, +1, \ldots, +11\}$, where the integer value represents the number of semitones from the previous note. Using the interval representation, the calculation of entropy and conditional entropy are the same as with scale degrees, except that the maximum entropy value is now $H_{max} = log_2 23 \approx 4.52$.

Results

ZEROTH-ORDER INFORMATION DISTRIBUTION OF SCALE DEGREES

We first examine the asymmetry in the zeroth-order information distribution of phonemes/scale degrees in words/musical segments. Figure 2 shows the zeroth-order information distribution (according to normalized entropy) within musical segments (using the above segmentation algorithms on the Essen and Wikifonia dataset, and scale degree representation) and words (from Yannakoudamis and Hutton's, 1992, data) of different lengths. We plotted the distributions using both the absolute position of the notes, and the normalized position, which is relative to the length of the segment/word. The plots show the overall average entropy (dashed black line), as well as the average entropy at each position (solid black line), which is calculated by taking the average over regularly-spaced bins along the x-axis.

To examine the asymmetry in the shape of the information distribution, we compared the average normalized entropy in four subsegments of the musical segments/words: the first note/letter, last note/letter, left half excluding first note/letter (denoted as left exclusive), and right-half excluding last note/letter (denoted as right exclusive). Figure 3 shows the comparisons over the two music datasets and words. For words, the information content has a "cliff" shape, $F(3, 8) = 75.31$, $p < .001$, $\eta_p^2 = .90$. Specifically, the information content of the last letter is significantly less than that of the other subsegments; last vs. first: $t(8) = 11.07$, $p < .001$; last vs. left exclusive: $t(8) = 7.81$, $p < .001$; last vs. right exclusive: $t(8) = 18.24$, $p < 0.001$, whereas there is no difference in information content between the other three

[3] Note that Yannakoudakis and Hutton (1992) did not report the number of words used to calculate the entropy distribution in each word length condition.

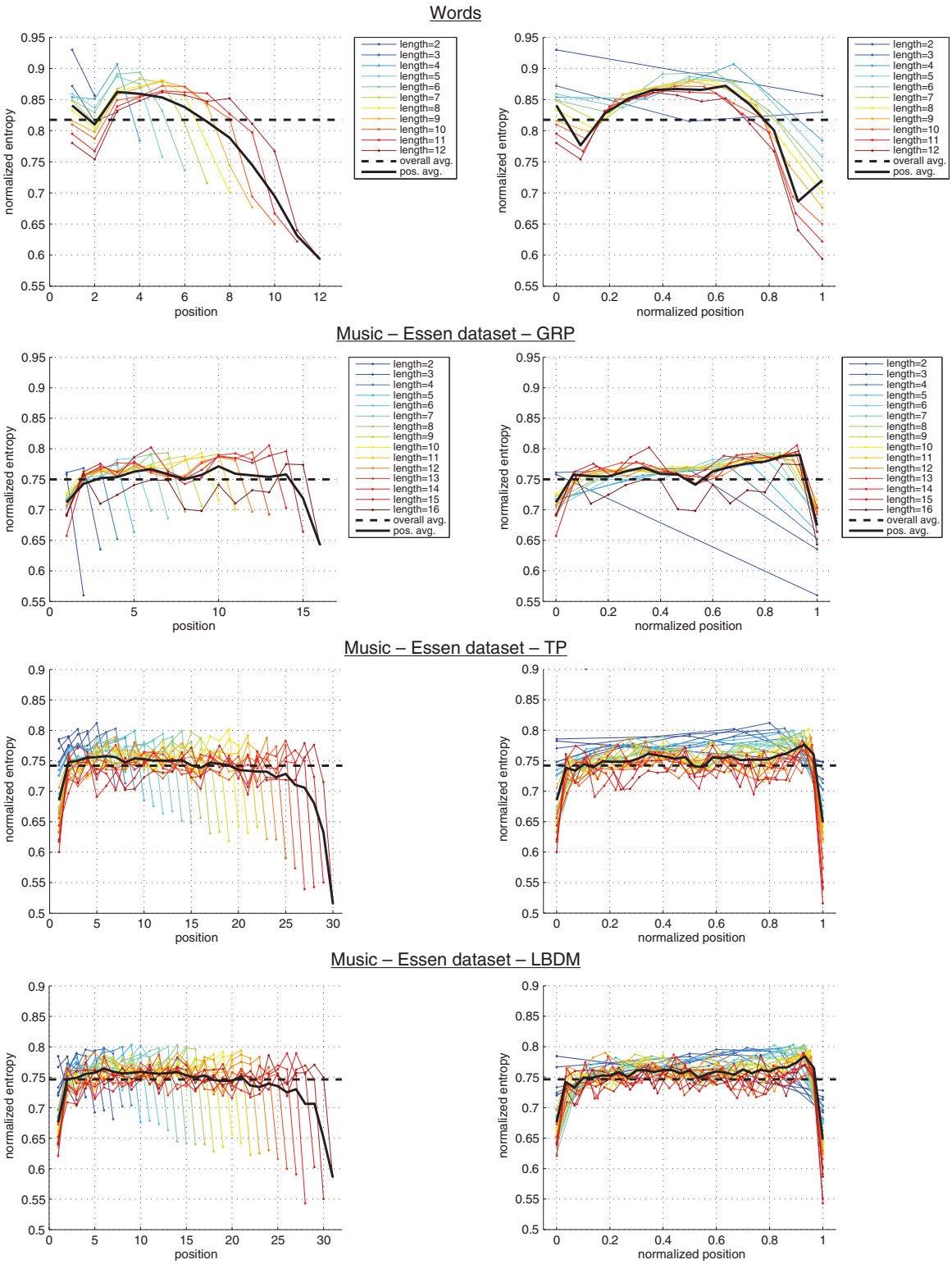

FIGURE 2. Entropy distribution of phonemes/scale degrees in words/musical segments of different length, using (left) absolute positions and (right) normalized positions. For music datasets, different music segmentation methods are presented in each row: Temporal Proximity (TP), Local Boundary Detection Model (LBDM), Grouper (GRP), Information Dynamics of Music (IDyOM), and human annotations (H). Each gray-level represents the distribution for phrases of a particular length. Dashed black line is the overall average entropy. Solid dashed line is the positional average entropy.

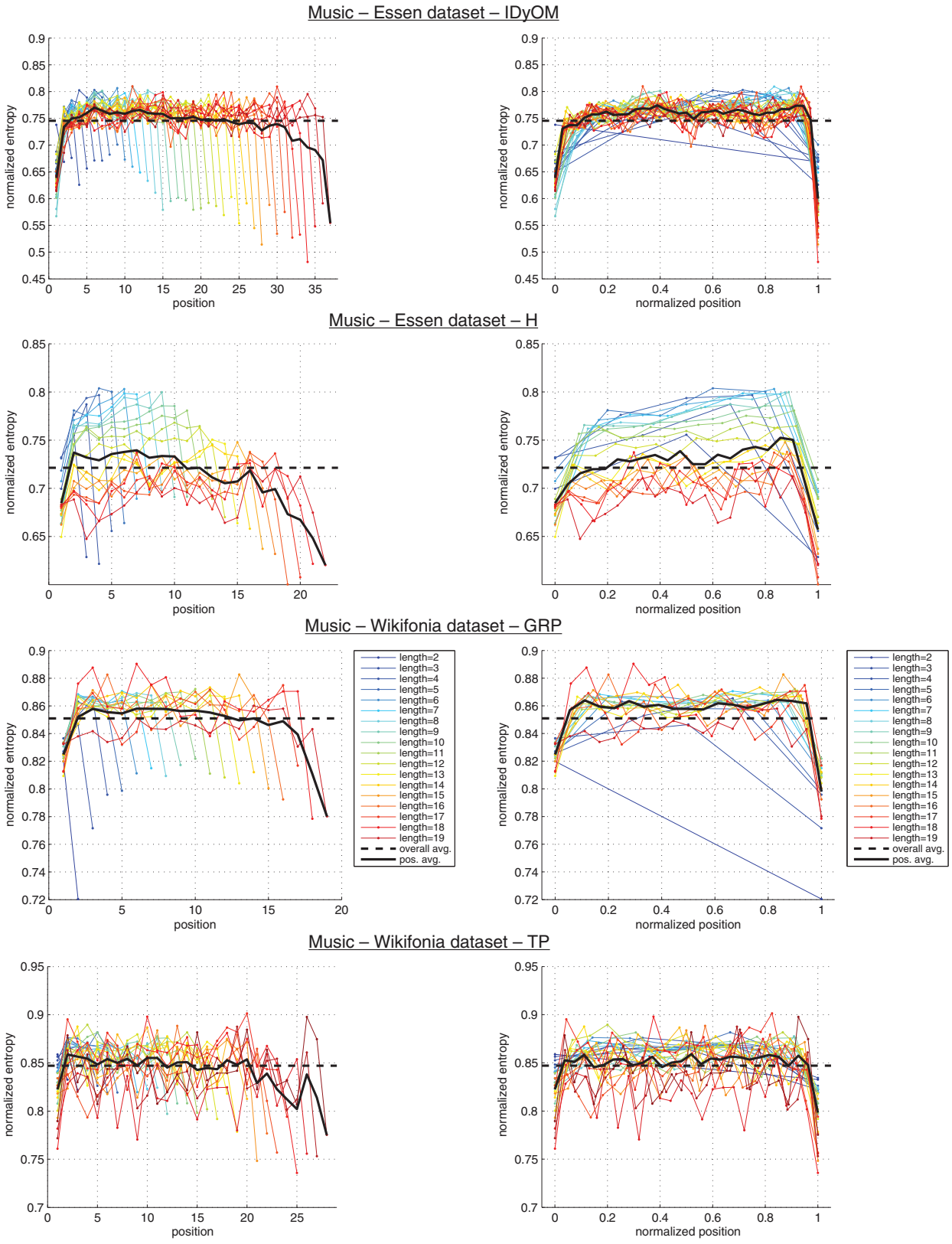

FIGURE 2. [Continued]

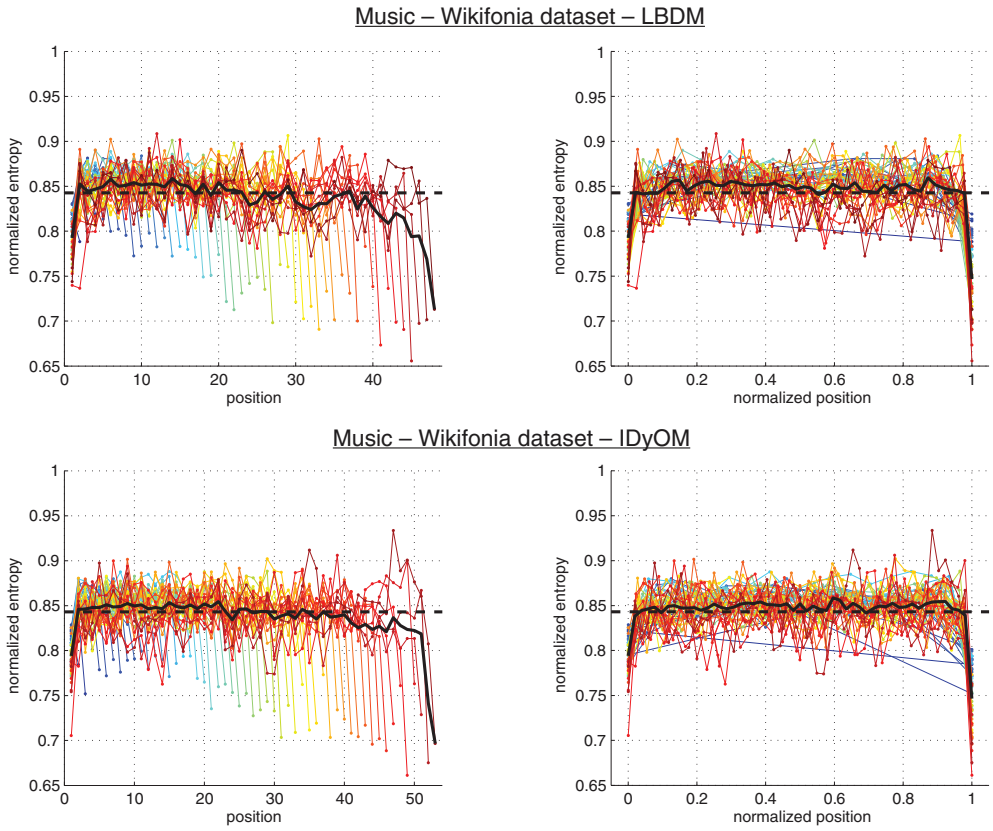

FIGURE 2. [Continued]

subsegments (first, left exclusive, and right exclusive). For musical segments extracted from the Essen dataset, the information content follows an asymmetric inverted U shape; GRP: $F(3, 12) = 114.12$, $p < .001$, $\eta_p^2 = .91$; TP: $F(3, 26) = 157.14$, $p < .001$, $\eta_p^2 = .86$; LBDM: $F(3, 27) = 269.07$, $p < .001$, $\eta_p^2 = .91$; IDyOM: $F(3, 33) = 342.51$, $p < .001$, $\eta_p^2 = .91$; H: $F(3, 18) = 85.87$, $p < .001$, $\eta_p^2 = .83$. The information content increases in the first three subsegments; e.g., Essen H, first to left exclusive: $t(18) = -7.82$, $p < .001$; left exclusive to right exclusive: $t(18) = -7.84$, $p < .001$, and then the information content of the last note drops to below that of the first note; e.g., Essen H: $t(18) = 2.48$, $p = .02$. This shape of the entropy distribution is consistent regardless of the segmentation method used to obtain the musical segments from Essen (see Figure 3a). For musical segments extracted from the Wikifonia dataset, the information distribution also follows an asymmetric inverted U shape; GRP: $F(3, 15) = 235.44$, $p < .001$, $\eta_p^2 = .94$; TP: $F(3, 24) = 124.61$, $p < .001$, $\eta_p^2 = .84$; LBDM: $F(3, 44) = 361.24$, $p < .001$, $\eta_p^2 = .89$; IDyOM: $F(3, 49) = 368.79$, $p < .001$, $\eta_p^2 = .88$, but with one key difference: the information content of the left exclusive and right exclusive subsegments are not different, i.e., the information content of the middle notes is flat. Specifically, the information content increases from the first note to the left exclusive half; e.g., Wiki GRP: $t(15) = -12.61$, $p < .001$, remains the same in the right exclusive half, and then the information content of the last note drops below that of the first note, $t(15) = 6.90$, $p < .001$. Again, this shape of the information distribution is consistent regardless of the segmentation method used to extract the musical segments (see Figure 3b). In both musical segments and words, the first note/letter has higher information content (higher entropy) than the last note/letter; words: $t(8) = 11.07$, $p < .001$; Essen-GRP: $t(12) = 2.41$, $p = .03$; Essen-TP: $t(26) = 5.92$, $p < .001$; Essen-LBDM: $t(27) = 4.47$, $p < .001$; Essen-IDyOM: $t(33) = 4.27$, $p < .001$; Essen-H: $t(18) = 2.48$, $p = .02$; Wiki-GRP: $t(15) = 6.90$, $p < .001$; Wiki-TP: $t(24) = 5.76$, $p < .001$; Wiki-LBDM: $t(44) = 9.90$, $p < .001$; Wiki-IDyOM: $t(49) = 9.59$, $p < .001$. However, words have a larger difference in information content between the first and last letters than musical segments; words vs.

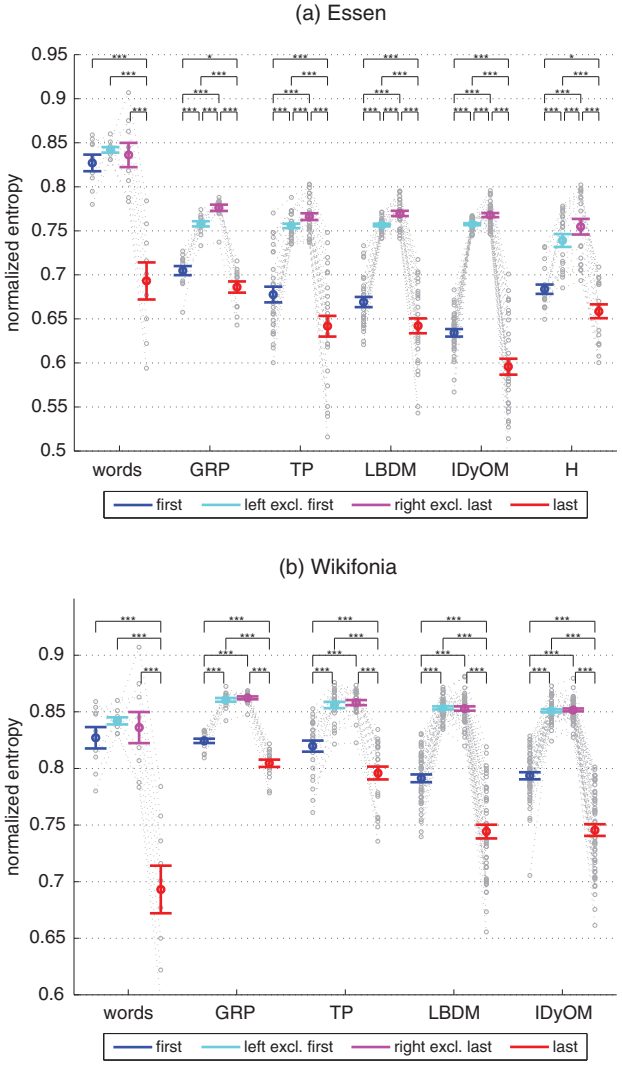

FIGURE 3. Comparison of average normalized entropy of phonemes/scale degrees in the first, left half excluding first, right half excluding last, and last positions of words/musical segments. Brackets at the top indicate significant differences between pairs (*$p < .05$, **$p < .01$, ***$p < .001$).

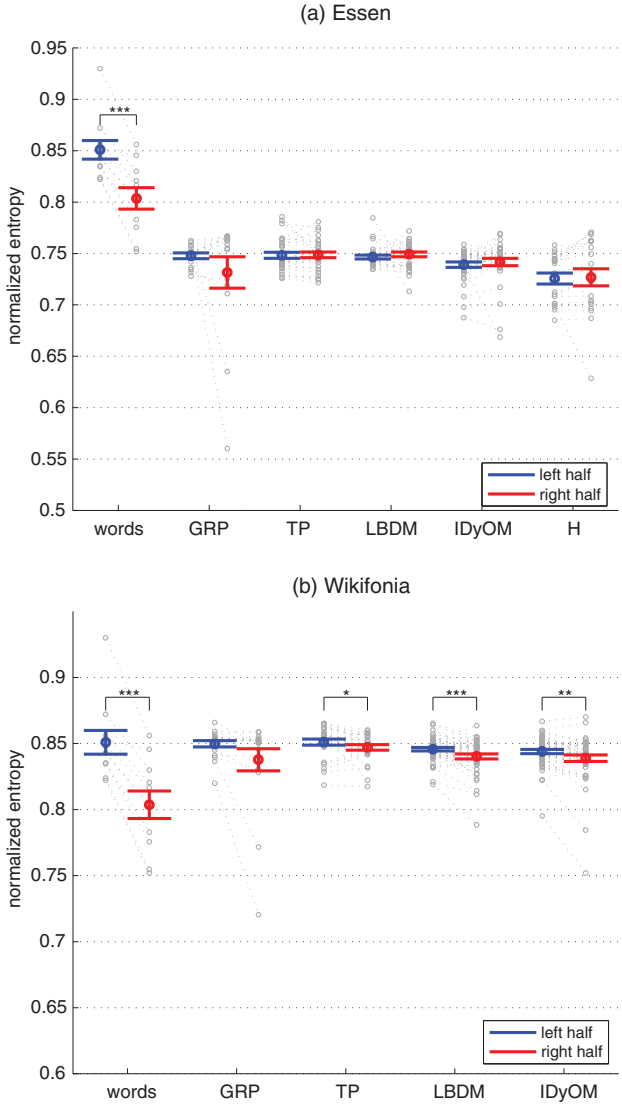

FIGURE 4. Comparison of average normalized entropy of phonemes/scale degrees in the left half and right half of words and musical segments (*$p < .05$, **$p < .01$, ***$p < .001$).

Essen-GRP: $t(20) = 8.40$, $p < .001$; vs. Essen-TP: $t(34) = 7.77$, $p < .001$; vs. Essen-LBDM: $t(35) = 8.45$, $p < .001$; vs. Essen-IDyOM: $t(41) = 5.12$, $p < .001$; vs. Essen-H: $t(26) = 6.43$, $p < .001$; vs. Wikifonia-GRP: $t(23) = 11.74$, $p < .001$; vs. Wikifonia-TP: $t(32) = 11.12$, $p < .001$; vs. Wikifonia-LBDM: $t(52) = 7.30$, $p < .001$; vs. Wikifonia-IDyOM: $t(57) = 6.70$, $p < .001$.

To further examine the asymmetry in the information distribution within words and musical segments, we compare the average normalized entropy in the beginning and ending halves (left and right) of words and musical segments in Figure 4. In words, the beginning half has higher information content than the ending half, $t(10) = 7.80$, $p < .001$. However, the left exclusive and right exclusive halves do not have a significant difference, $t(8) = 0.50$, $p = .63$, which suggests that the asymmetric information distribution in words is mainly due to the difference in entropy between the first and last letters. For Wikifonia musical segments, the left half has higher information content than the right half for TP, LBDM, and IDyOM; TP: $t(26) = 2.06$, $p < .05$; LBDM: $t(46) = 3.77$, $p < .001$; IDyOM: $t(51) = 3.47$, $p = .001$, whereas for GRP, the difference between left and right halves did not reach significance, $t(17) = 1.91$,

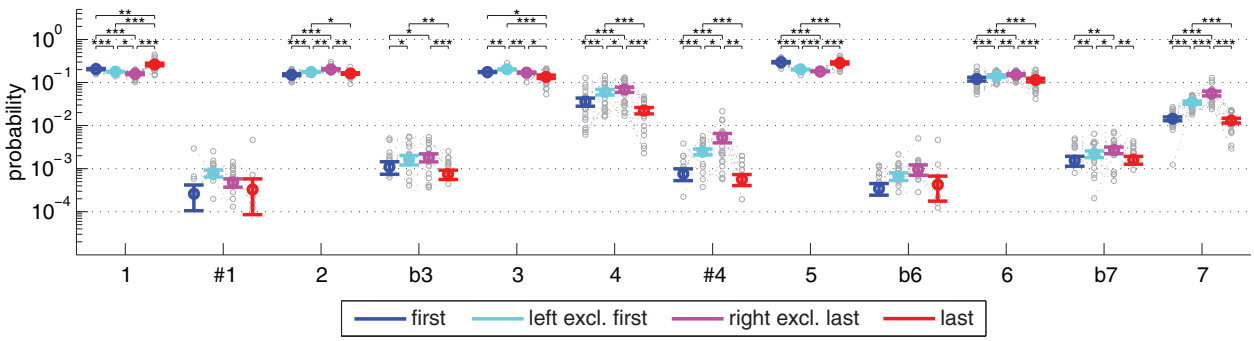

FIGURE 5. Probabilities of scale degrees for the first note, left half excluding first note, right half excluding last note, and last note. Probabilities were calculated from the musical segments extracted using human annotations of Essen (*$p < .05$, **$p < .01$, ***$p < .001$).

$p = .07$. Since the left exclusive and right exclusive halves of the musical segments did not have a significant difference in entropy in Wikifonia, this again suggests the asymmetric information distribution is due to the difference in entropy (information content) in the first and last notes, similar to words. However, words have more asymmetric left and right halves than musical segments; words vs. Wikifonia-GRP: $t(27) = 3.75$, $p < .001$; vs. Wikifonia-TP: $t(36) = 8.90$, $p < .001$; vs. Wikifonia-LBDM: $t(56) = 10.08$, $p < .001$; vs. Wikifonia-IDyOM: $t(61) = 10.18$, $p < .001$. Finally, for musical segments extracted from the Essen dataset, there is no significant difference in entropy between the left and right halves (see Figure 4a). Since the information distribution in the Essen dataset has an asymmetric inverted U shape (see Figure 3a), this suggests that the decrease in entropy between the first and last notes is the same magnitude as the increase in entropy between the left exclusive and right exclusive halves.

SCALE DEGREE DISTRIBUTIONS FOR ZEROTH-ORDER MODEL

The analysis in the previous section indicates that the information content of musical segments follows an inverted U shape. We next examine the distributions of scale degrees within musical segments. The probabilities of each scale degree occurring in the four subsegments (first note, left exclusive, right exclusive, and last note) for the Essen musical segments from human annotations (H) are shown in Figure 5.[4]

There are three main observations. First, the probability profiles of scale degrees 1 and 5 follow an asymmetric U shape, where these scale degrees occur less frequently in the middle of the segment and more frequently in the first and last note; 1: $F(3, 18) = 15.60$, $p < .001$, $\eta_p^2 = .46$; 5: $F(3, 18) = 28.43$, $p < .001$, $\eta_p^2 = .61$. Scale degree 1 is more likely to occur as the last note than the first note, $t(18) = -3.12$, $p = .01$, whereas in contrast, there is no difference in likelihood of scale degree 5 appearing in the first or last note, $t(18) = 0.44$, $p = .67$. In the middle of the musical segment, both scale degrees 1 and 5 are more likely to occur in the left exclusive half than in the right exclusive; 1: $t(18) = 2.67$, $p = .02$; 5: $t(18) = 8.51$, $p < .001$. Second, a large number of other scale degrees (2, b3, 4, #4, 6, b7, 7) have an asymmetric inverted U shape; 2: $F(3, 18) = 13.99$, $p < .001$, $\eta_p^2 = .44$; b3: $F(3, 18) = 8.02$, $p < .001$, $\eta_p^2 = .31$; 4: $F(3, 18) = 23.86$, $p < .001$, $\eta_p^2 = .57$; #4: $F(3, 18) = 12.68$, $p < .001$, $\eta_p^2 = .41$; 6: $F(3, 18) = 19.42$, $p < .001$, $\eta_p^2 = .52$; b7: $F(3, 18) = 6.35$, $p < .001$, $\eta_p^2 = .26$; 7: $F(3, 18) = 42.58$, $p < .001$, $\eta_p^2 = .70$. The probability of these scale degrees increases from the first note to the left exclusive subsegment; 2: $t(18) = -4.82$, $p < .001$; b3: $t(18) = -2.27$, $p = .04$; 4: $t(18) = -4.08$, $p < .001$; #4: $t(18) = -4.60$, $p < .001$; 6: $t(18) = -4.70$, $p < .001$; b7: $t(18) = -3.06$, $p = .007$; 7: $t(18) = -8.94$, $p < .001$, and further increases in the right exclusive; 2: $t(18) = -3.65$, $p = .002$; 4: $t(18) = -2.25$, $p = .04$; #4: $t(18) = -2.38$, $p = .03$; 6: $t(18) = -3.77$, $p = .001$; b7: $t(18) = -2.14$, $p = .05$; 7: $t(18) = -3.99$, $p < .001$. Then, the probability of the scale degree in the last note decreases to the same level of the first note, i.e., there was no significant difference in the probability of the scale degree occurring in the first or last note; 2: $t(18) = -1.38$, $p = .19$; b3: $t(18) = 1.41$, $p = .18$; 4: $t(18) = 1.98$, $p = .06$; #4: $t(18) = 0.83$, $p = .42$; 6: $t(18) = 0.67$, $p = .51$; b7: $t(18) = -0.17$, $p = .87$; 7: $t(18) = 0.94$, $p = .36$. Third, the probability profile of scale degree 3 also has an inverted U shape, but in contrast to others, it is more likely to occur in the

[4] Similar results were obtained from musical segments of the automatic methods.

left exclusive half than the right exclusive, $t(18) = 3.44$, $p = .003$, and is more likely to occur in the first note than the last note, $t(18) = 2.66$, $p = .02$.

The analysis of the zeroth-order scale degree probabilities suggests an explanation for the information distribution in musical segments. The increased entropy in the middle of the musical segment is due to the increased likelihood of 8 scale degrees (2, b3, 3, 4, #4, 6, b7, 7), which correspond to notes in various common scales (e.g., major, Dorian, Lydian, Mixolydian). Within the middle, the increase in probability of the 8 scale degrees in the right exclusive subsegment, along with a corresponding decrease in probability of scale degrees 1 and 5, leads to higher entropy in the right exclusive half. In contrast, the first and last note have lower entropy than the middle because of the increased probability of scale degrees 1 and 5 as a first and last note, with corresponding decreased likelihoods of all other scale degrees. Since the probability of scale degree 5 is similar for the first and last note, the difference in entropy between the first and last notes is mainly due to the increased probability of scale degree 1 as the last note (and correspondingly, a decreased probability of scale degree 3).

FIRST-ORDER INFORMATION DISTRIBUTION OF SCALE DEGREES

We next examine the first-order information distribution of scale degrees in musical segments. Figure 6 shows the average normalized first-order conditional entropy using absolute note positions and normalized positions (similar to Figure 2) for musical segments from Essen and Wikifonia.

To examine the asymmetry in the 1st order information distribution, we again compare the average conditional entropy in four subsegments (first, left exclusive, right exclusive, last), which is presented in Figure 7.

For Essen musical segments, the first-order information distribution follows a "cliff" shape; GRP: $F(3, 11) = 231.99$, $p < .001$, $\eta_p^2 = .96$; TP: $F(3, 25) = 258.75$, $p < .001$, $\eta_p^2 = .91$; LBDM: $F(3, 26) = 452.11$, $p < .001$, $\eta_p^2 = .95$; IDyOM: $F(3, 32) = 563.72$, $p < .001$, $\eta_p^2 = .95$; H: $F(3, 17) = 226.07$, $p < .001$, $\eta_p^2 = .93$. Specifically, the last note has significantly lower entropy than the rest of the musical segments; e.g., for H, last vs. first: $t(17) = 16.80$, $p < .001$; last vs. left exclusive: $t(17) = 17.18$, $p < .001$; last vs. right exclusive: $t(17) = 2.57$, $p = .02$. This observation is consistent for all segmentation methods on Essen (see Figure 7). There are also some differences in conditional entropy between the first, left exclusive and right exclusive halves, but it depends on the segmentation method used; GRP: first vs. left exclusive: $t(11) = 2.34$, $p = .04$; TP: left exclusive vs. right exclusive: $t(25) = -2.51$, $p = .02$; IDyOM: left exclusive vs. right exclusive: $t(32) = 2.58$, $p = .01$; H: first vs. right exclusive: $t(17) = 2.57$, $p = .02$. However, these differences are small in magnitude when compared with the decrease in conditional entropy of the last note. For the musical segments extracted from Wikifonia, the 1st order information distribution follows an inverted U shape; GRP: $F(3, 14) = 50.21$, $p < .001$, $\eta_p^2 = .78$; TP: $F(3, 23) = 17.57$, $p < .001$, $\eta_p^2 = .43$; LBDM: $F(3, 43) = 165.46$, $p < .001$, $\eta_p^2 = .79$; IDyOM: $F(3, 48) = 272.93$, $p < .001$, $\eta_p^2 = .85$, with small increases from the first to right exclusive regions; GRP: $t(14) = -5.85$, $p < .001$; TP: $t(23) = -4.66$, $p < .001$; LBDM: $t(43) = -5.93$, $p < .001$; IDyOM: $t(48) = -4.24$, $p < .001$, before dropping significantly in the last note; GRP: $t(14) = 4.40$, $p < .001$; TP: $t(23) = 2.12$, $p < .05$; LBDM: $t(43) = 9.75$, $p < .001$; IDyOM: $t(48) = 15.68$, $p < .001$.

Finally, we examine the asymmetry in the 1st order information distribution by comparing the left and right halves of the musical segments (Figure 8). On Essen, the left half has higher 1st order information content than the right half for all segmentation methods; GRP: $t(13) = 3.95$, $p = .002$; TP: $t(27) = 13.16$, $p < .001$; LBDM: $t(28) = 8.04$, $p < .001$; IDyOM: $t(34) = 7.97$, $p < .001$; H: $t(19) = 4.65$, $p < .001$. On Wikifonia, the results are mixed. For LBDM and IDyOM, the left half has higher 1st order information content than the right half; LBDM: $t(45) = 3.05$, $p = .004$; IDyOM: $t(50) = 6.27$, $p < .001$. On the opposite, for TP segments, the left half has lower information content than the right half, $t(25) = -2.33$, $p = .03$. Finally, there is no significant difference between left and right halves for GRP segments.

INFORMATION DISTRIBUTIONS OF PITCH INTERVALS

Here we examine the asymmetry in the shape of the information distribution of pitch intervals in musical segments. Figure 9 shows the zeroth-order information distribution of pitch intervals in the four subsegments (first, left exclusive, right exclusive, last) and left/right halves of musical segments.

On Essen, the information distribution of pitch intervals in musical segments by LBDM, IDyOM, and Humans have an inverted U shape; LBDM: $F(3, 26) = 78.30$ $p < .001$, $\eta_p^2 = .75$; IDyOM: $F(3, 32) = 126.63$, $p < .001$, $\eta_p^2 = .80$; H: $F(3, 17) = 35.67$, $p < .001$, $\eta_p^2 = .68$, while those by GRP and TP have a "cliff" shape; GRP: $F(3, 11) = 6.16$, $p = .002$, $\eta_p^2 = .36$; TP: $F(3, 25) = 35.13$, $p < .001$, $\eta_p^2 = .58$. For all 5 sets of musical segments, the last pitch interval had lower entropy than the other 3 subsegments; e.g., for H, first vs. last:

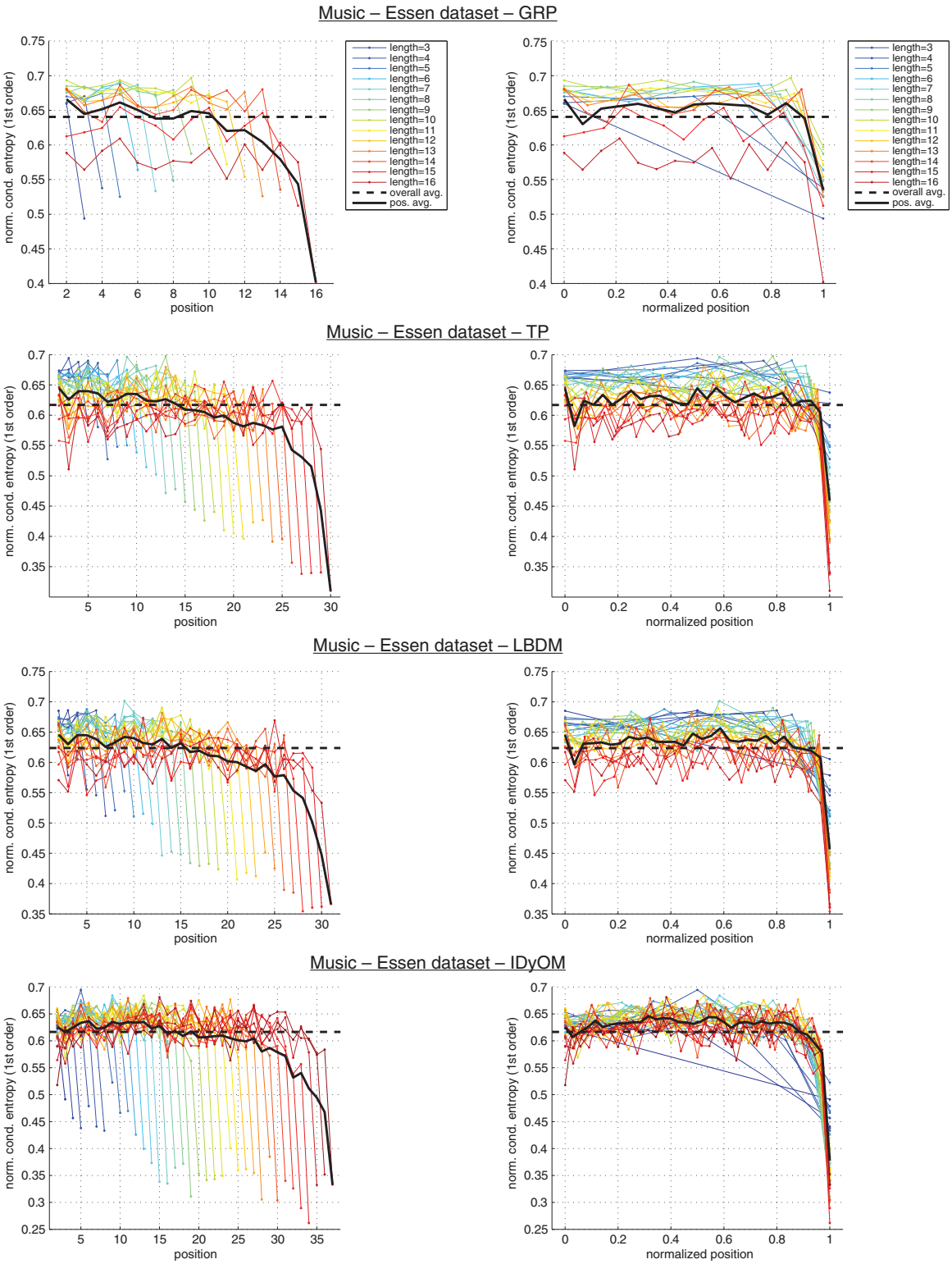

FIGURE 6. First-order conditional entropy distribution of scale degrees in musical segments of different length, using (left) absolute positions and (right) normalized positions. Different music segmentation methods are presented in each row: Temporal Proximity (TP), Local Boundary Detection Model (LBDM), Grouper (GRP), Information Dynamics of Music (IDyOM), and human annotations (H). Each gray-level represents the distribution for phrases of a particular length. Dashed black line is the overall average entropy. Solid dashed line is the positional average entropy.

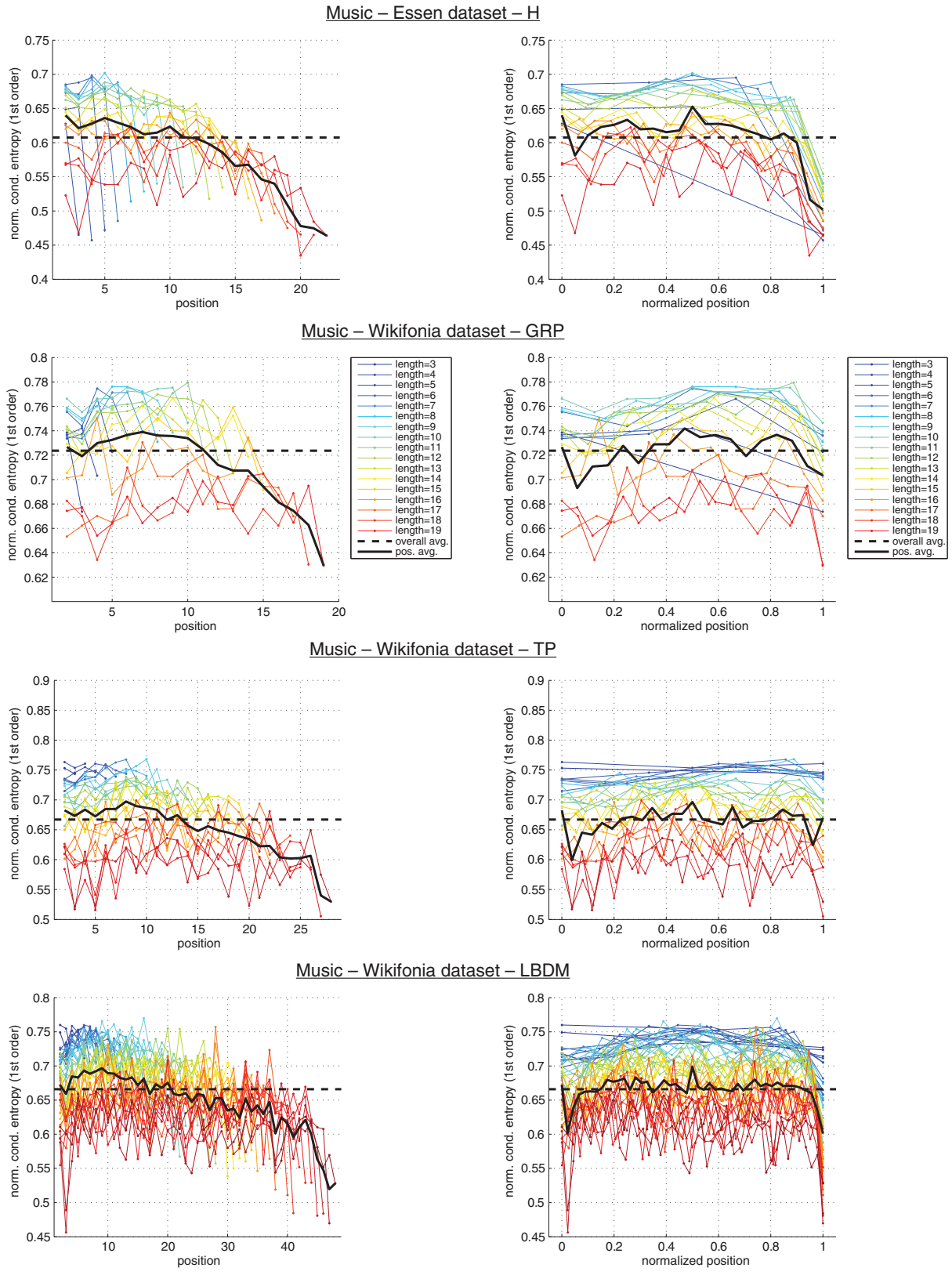

FIGURE 6. [Continued]

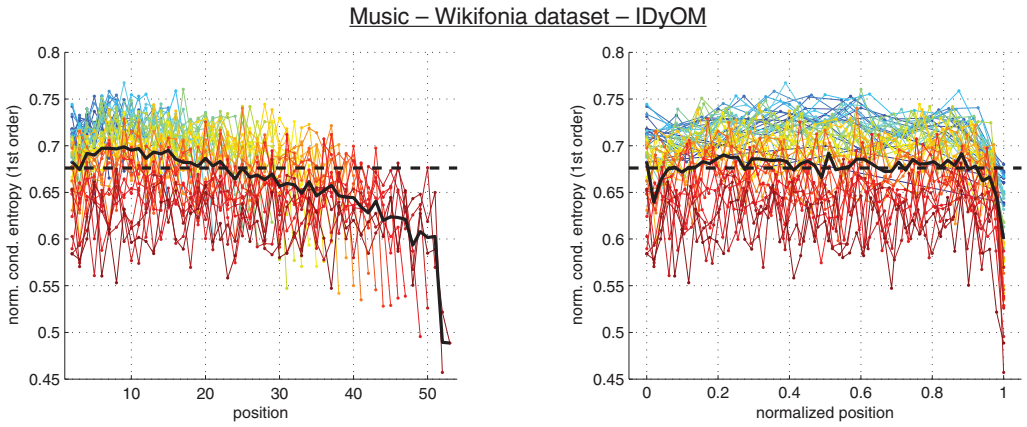

FIGURE 6. [Continued]

$t(17) = 5.36$, $p < .001$; left exclusive vs. last: $t(17) = 9.65$, $p < .001$; right exclusive vs. last: $t(17) = 10.05$, $p < .001$. This suggests that the resolution of the penultimate note to the last note is more predictable (less entropy) than other notes in the musical segments in Essen. On Wikifonia, the information distributions also have an inverted U for LBDM and IDyOM; LBDM: $F(3, 43) = 62.53$, $p < .001$, $\eta_p^2 = .59$; IDyOM: $F(3, 48) = 32.02$, $p < .001$, $\eta_p^2 = .40$, but with the last note having higher entropy than the first note; e.g., LBDM, first vs. last: $t(43) = -6.65$, $p < 0.001$, which is in contrast to Essen. For GRP and TP, the information distributions in Wikifonia have an increasing trend; GRP: $F(3, 14) = 69.23$, $p < .001$, $\eta_p^2 = .83$; TP: $F(3, 23) = 94.65$, $p < .001$, $\eta_p^2 = .81$, with the last note having the highest entropy; e.g., for GRP, first vs. last: $t(14) = -9.43$, $p < .001$; left exclusive vs. last: $t(14) = -9.06$, $p < .001$; right exclusive vs. last: $t(14) = -3.76$, $p = .002$. These results suggest an interesting phenomenon in Wikifonia: although the final notes (scale degrees) are most predictable (lowest entropy, see Figure 3b) in both Wikifonia and Essen, the resolution from the penultimate note to the final note is less predictable than in Essen. Indeed, the average normalized entropy of scale degrees in musical segments from Wikifonia is higher than that in Essen; GRP: 0.85 ± 0.02 vs. 0.74 ± 0.03, $t(31) = -12.22$, $p < .001$; TP: 0.85 ± 0.01 vs. 0.75 ± 0.01, $t(54) = -29.33$, $p < .001$; LBDM: 0.84 ± 0.01 vs. 0.75 ± 0.01, $t(75) = -46.05$, $p < .001$; IDyOM: 0.84 ± 0.01 vs. 0.74 ± 0.01, $t(86) = -37.24$, $p < .001$.

Next we examine at the pitch interval probabilities. Figure 10 shows the probabilities for each pitch interval within the four subsegments of a musical segment using human-annotated musical segments.

The probability decreases for larger pitch intervals, reflecting the well-known property of pitch proximity (Narmour, 1990; Temperley, 2014; von Hippel, 2000). Further examining the pitch interval probabilities within musical segments can help to explain the reduction in entropy of the last pitch interval. In particular, a small set of pitch intervals (-9, -7, -5, -4, -2, $+1$) have increased probability at the end of the segments (see Figure 10). In contrast, the other pitch intervals (-10, -8, -3, -1, 0, $+2$, $+3$, $+4$, $+5$, $+7$, $+8$, $+9$) have decreased probability at the end of the segments (see Figure 10). As a result of these two trends, the entropy of the last pitch interval decreases (see Figure 9a). In addition, looking at the trend within musical segments, many negative pitch intervals (-9, -7, -5, -4, -3, -2, $+1$) have probabilities that increase towards the segment endings (see Figure 10). In contrast, positive pitch intervals (0, $+3$, $+4$, $+5$, $+7$, $+9$) have probabilities decreasing towards the segment endings (see Figure 10). These phenomena are consistent with "melodic arches" observed in the literature (Huron, 2006): the overall pitch contour tends to rise and then fall over the course of a melodic phrase. In other words, falling pitch intervals are more likely in the end of the segments, while rising pitch intervals are more likely in the beginning of the segments.

Finally, we examine the asymmetry in the 1st order information distribution of pitch intervals. Figure 11 shows the average conditional entropy in the four subsegments in Essen and Wikifonia.

On Essen, the first-order conditional entropy has an inverted U shape; -10: $F(3, 17) = 6.08$, $p = .001$, $\eta_p^2 = .26$; -9: $F(3, 17) = 14.55$, $p < .001$, $\eta_p^2 = .46$; -8: $F(3, 17) = 12.43$, $p < .001$, $\eta_p^2 = .42$; -7: $F(3, 17) = 14.14$, $p < .001$, $\eta_p^2 = .45$; -6: $F(3, 17) = 5.52$, $p = .002$, $\eta_p^2 = .25$; -5: $F(3, 17) = 28.58$, $p < .001$, $\eta_p^2 = .63$; -4: $F(3, 17) = 20.88$, $p < .001$, $\eta_p^2 = .55$; -3: $F(3, 17) = 24.94$,

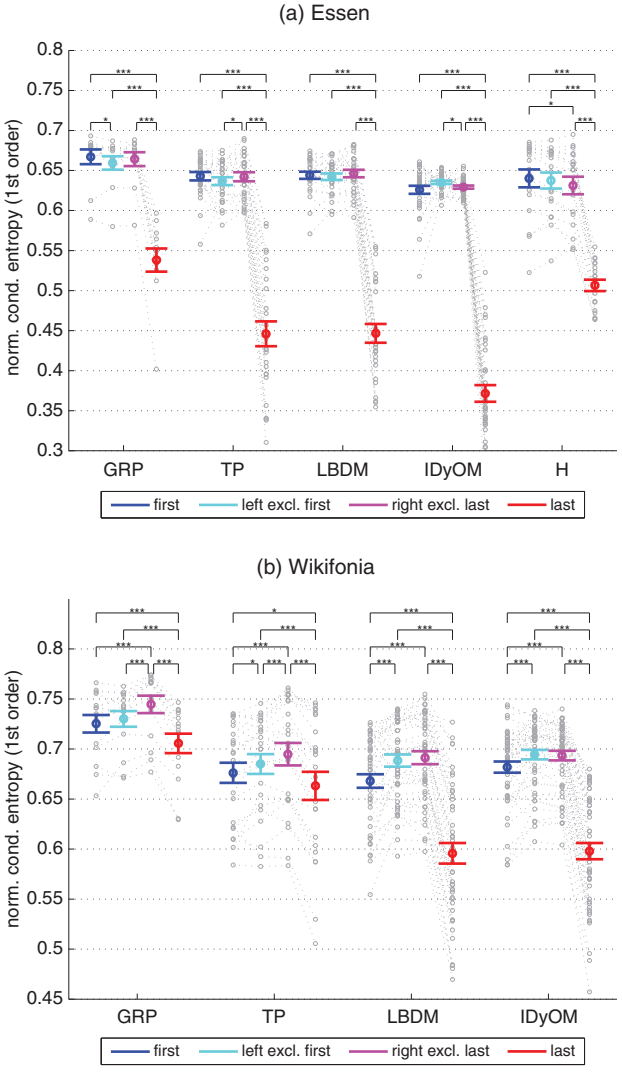

FIGURE 7. Comparison of the average normalized first-order conditional entropy of scale degrees in the first, left half excluding first, right half excluding last, and last notes of musical segments. Brackets at the top indicate significant differences between pairs ($p < .05$, $p < .01$, $p < .001$).

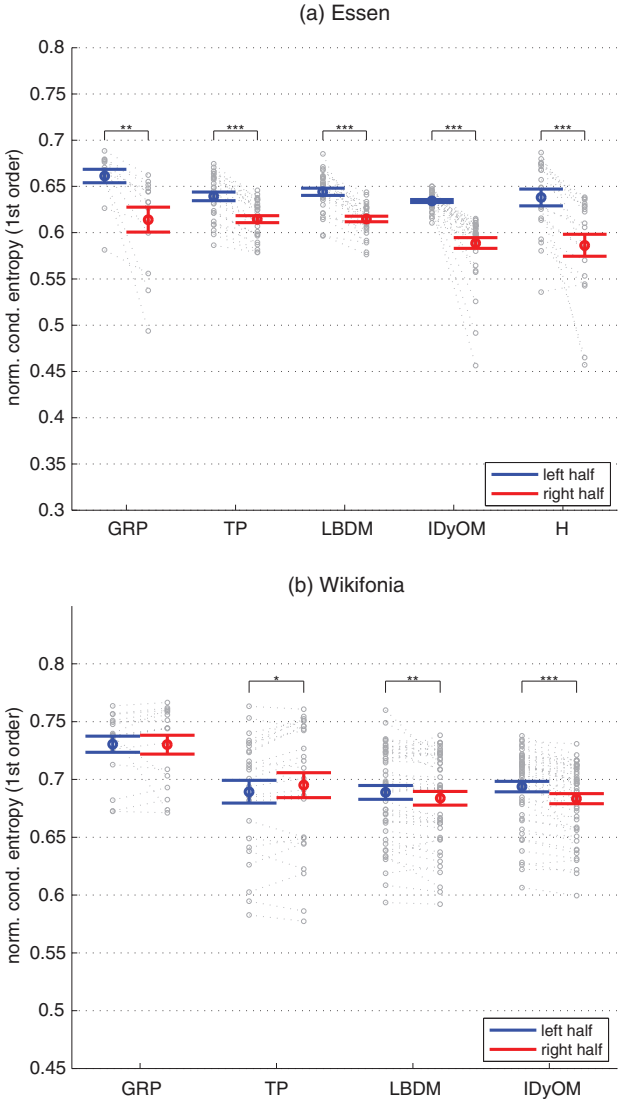

FIGURE 8. Comparison of average normalized first-order conditional entropy of scale degrees in the left and right halves of musical segments.

$p < .001$, $\eta_p^2 = .60$; −2: $F(3, 17) = 266.91$, $p < .001$, $\eta_p^2 = .94$; −1: $F(3, 17) = 10.75$, $p < .001$, $\eta_p^2 = .39$; 0: $F(3, 17) = 80.04$, $p < .001$, $\eta_p^2 = .83$; +1: $F(3, 17) = 13.23$, $p < .001$, $\eta_p^2 = .44$; +2: $F(3, 17) = 48.59$, $p < .001$, $\eta_p^2 = .74$; +3: $F(3, 17) = 19.91$, $p < .001$, $\eta_p^2 = .54$; +4: $F(3, 17) = 34.66$, $p < .001$, $\eta_p^2 = .67$; +5: $F(3, 17) = 45.38$, $p < .001$, $\eta_p^2 = .73$; +7: $F(3, 17) = 61.09$, $p < .001$, $\eta_p^2 = .78$; +8: $F(3, 17) = 12.92$, $p < .001$, $\eta_p^2 = .43$; +9: $F(3, 17) = 11.40$, $p < .001$, $\eta_p^2 = .40$, similar to the zeroth order entropy on Essen. In particular, for all segmentation methods, the normalized conditional entropy of the first and last notes are lower than the middle notes; e.g., for H, first vs. left exclusive: $t(16) = -7.78$, $p < .001$; first vs. right exclusive: $t(16) = -4.54$, $p < .001$; left exclusive vs. last: $t(16) = 20.15$, $p < .001$; right exclusive vs. last: $t(16) = 14.64$, $p < .001$, while the last note has lowest conditional entropy; for H, first vs. last: $t(16) = 5.71$, $p < .001$. In addition, the right-half of the musical segments have lower conditional entropy than the left half; e.g., for H, $t(18) = 6.04$, $p < .001$. On Wikifonia, the conditional entropy for pitch intervals also has similar shapes to the zeroth order entropy within a musical segment. In particular, for GRP and TP segmentations, the conditional entropy increases from the beginning to

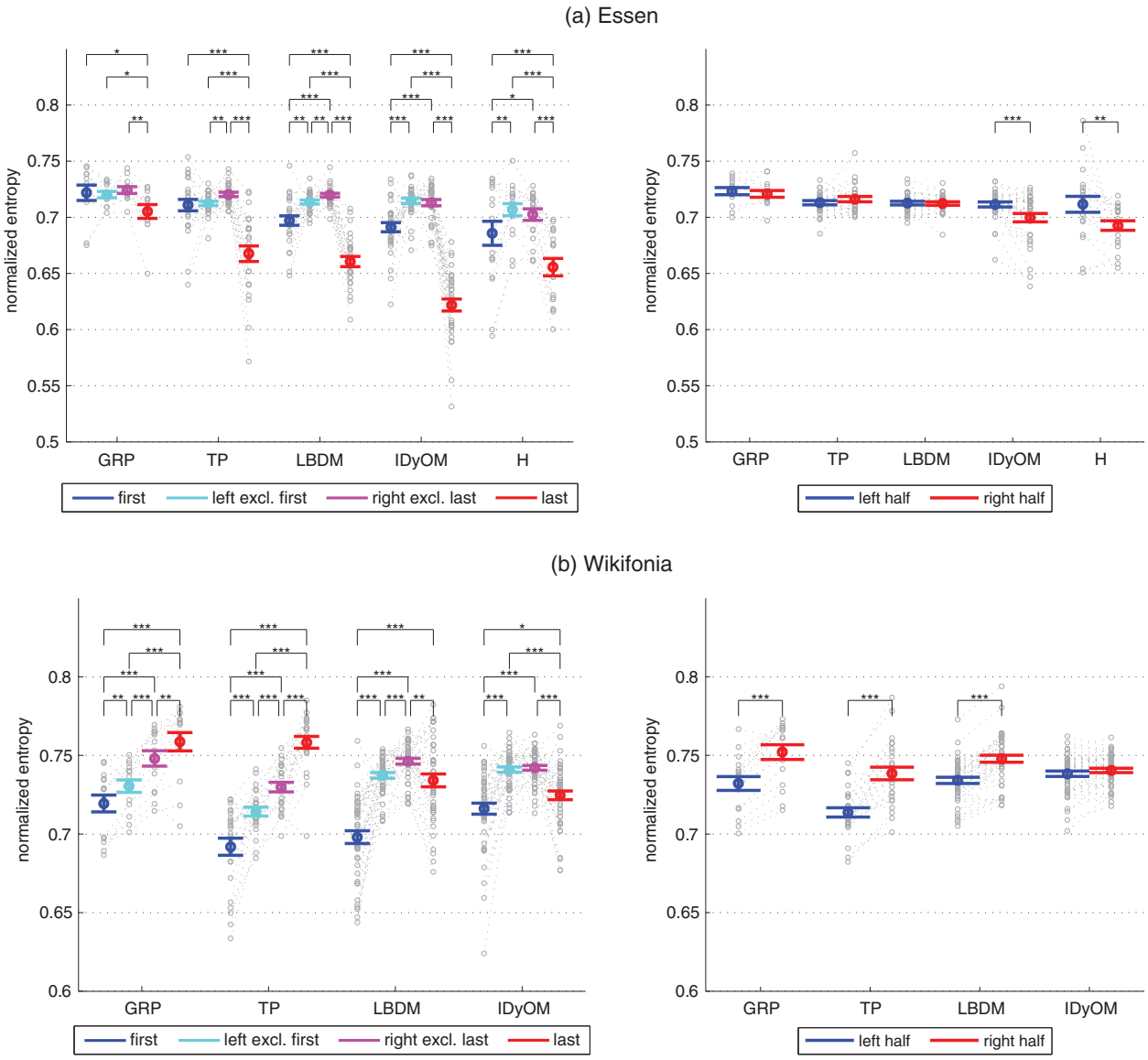

FIGURE 9. Comparison of average normalized entropy of pitch intervals in the first, left half excluding first, right half excluding last, and last positions of musical segments (left), and in the left half and right-half of musical segments (right). Brackets at the top indicate significant differences between pairs (*$p < .05$, **$p < .01$, ***$p < .001$).

the end of the segment; GRP: $F(3, 13) = 123.81$, $p < .001$, $\eta_p^2 = .91$; TP: $F(3, 22) = 148.88$, $p < .001$, $\eta_p^2 = .87$, while for LBDM and IDyOM, the conditional entropy has an inverted U shape; LBDM: $F(3, 42) = 169.88$, $p < .001$, $\eta_p^2 = .80$; IDyOM: $F(3, 47) = 97.26$, $p < .001$, $\eta_p^2 = .67$. On Wikifonia, the right half of the musical segments has higher 1st order conditional entropy than the left half for all segmentation methods; GRP: $t(15) = -6.97$, $p < .001$; TP: $t(24) = -5.83$, $p < .001$; LBDM: $t(44) = -5.45$, $p < .001$; IDyOM: $t(49) = -2.80$, $p = .007$.

Discussion

Here we investigated the information distribution within musical segments by analyzing the entropy at different locations of the segments obtained from two representative song datasets, Essen and Wikifonia, which predominantly contain folksongs and popular music, respectively. We used four computational methods to extract musical segments from the songs, and showed that these four methods roughly form different

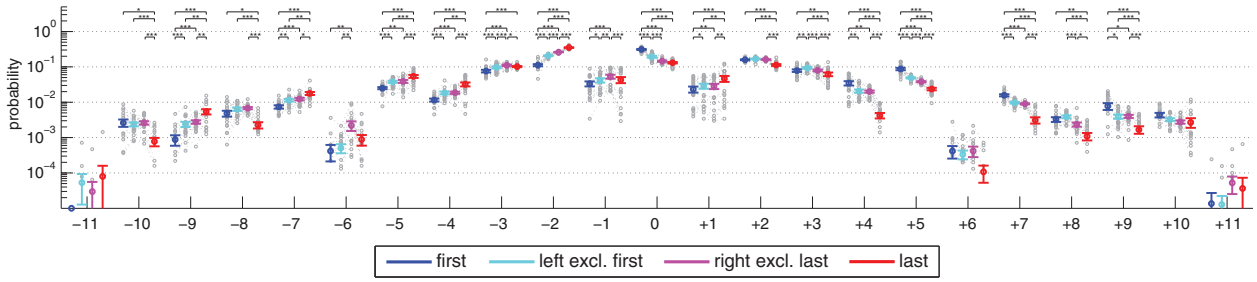

FIGURE 10. Probabilities of pitch intervals for the first note, left half excluding first note, right half excluding last note, and last note. Probabilities were calculated from the musical segments extracted using human annotations of Essen (*p < .05, **p < .01, ***p < .001).

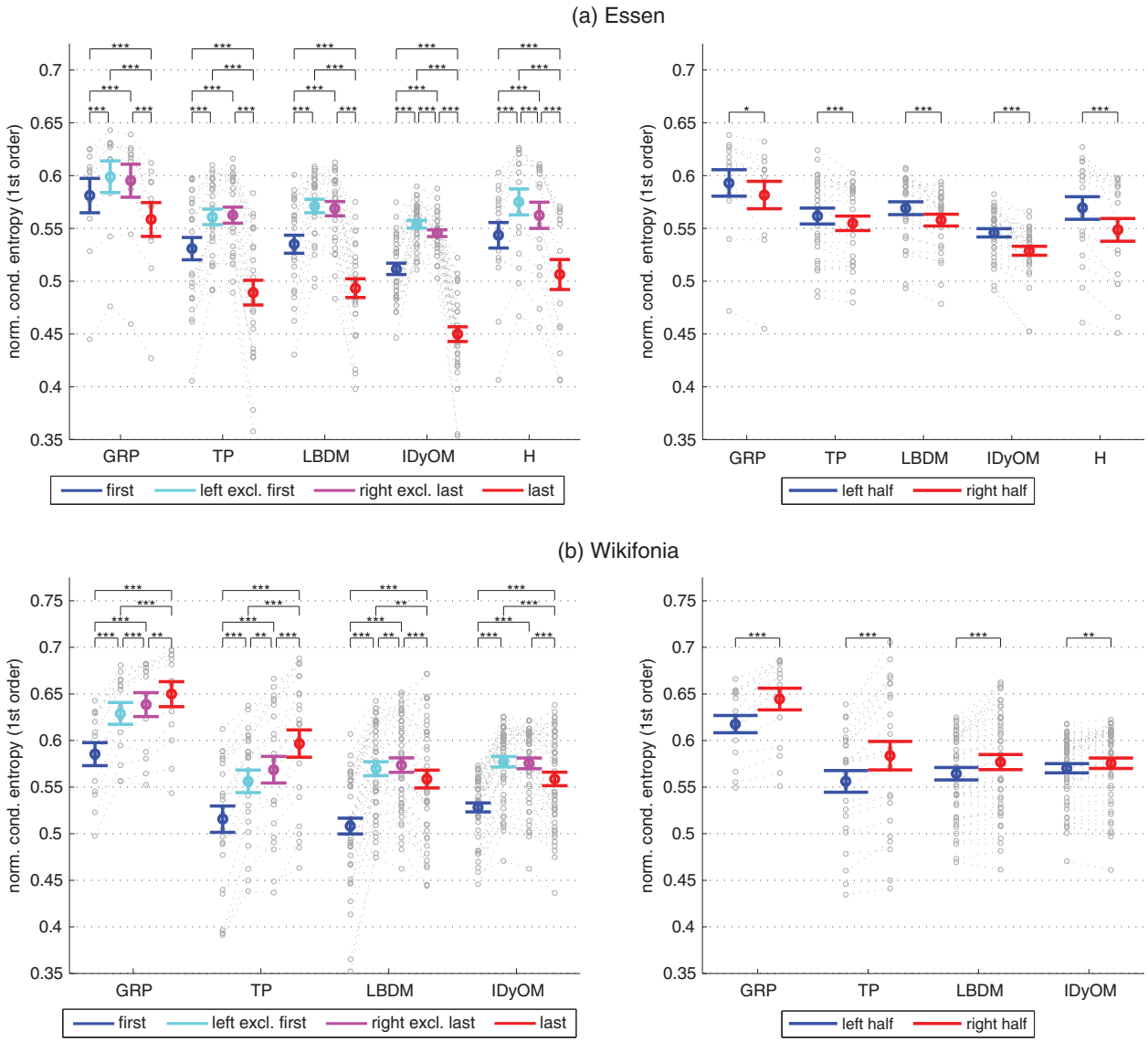

FIGURE 11. Comparison of average normalized first-order conditional entropy of pitch intervals in the first, left half excluding first, right half excluding last, and last positions of musical segments (left), and in the left half and right half of musical segments (right). Brackets at the top indicate significant differences between pairs (*p < .05, **p < .01, ***p < .001).

levels in the grouping structure hierarchy. The results showed that, regardless of the segmentation method used, the zeroth-order information distribution within musical segments exhibited an inverted U shape pattern, with the highest entropy in the middle of the segment, followed by the segment beginnings (first note); whereas the lowest entropy was observed in the segment endings, i.e., last note (Figure 3). The general shape of the inverted U was found in both datasets, but with a key difference in the middle of the segments – in Essen folksongs, higher entropy was observed in the right half of the middle than the left half, whereas in Wikifonia songs, both sides of the middle had the same entropy. The asymmetry in information distribution between segment beginnings and endings was similar to that observed in English words reported in the literature (e.g., Carstairs-McCarthy, 2002; Shillcock et al., 1996, 2000; Yannakoudakis & Hutton, 1992). In addition, the asymmetric information distribution within both musical segments and words was driven mainly by the entropy (information) difference between the first and the last notes/letters, as no significant asymmetry was observed when we excluded the first and the last notes/letters (Figure 3). Note however that there were some key differences between the information distributions within words and musical segments: (1) the asymmetry between the beginning and the ending halves was larger/more dramatic in words than in musical segments; and (2) the entropy at the word beginnings was around the average entropy, whereas the entropy at the musical segment beginnings was below the average entropy, with the highest entropy occurring in the middle of the segments, demonstrating an inverted U pattern (see Figure 2, 3 and 4).

The higher entropy at word beginnings than word endings has been argued to be related to a communicative pressure to express a maximum amount of information at word beginnings to allow a spoken word to be recognized as early as possible (e.g., Brysbaert & Nazir, 2005; Shillcock et al., 2000). In the case of music, this argument does not seem plausible; there is no musical "lexicon" of segments, analogous to a lexicon of words, in relation to which a musical pattern must be looked up in order to be understood. And in any case, as noted above, the information distributions in musical segments and words are rather different. This suggests that the asymmetry in the information distribution within musical segments may be due to factors different from those influencing that in words. In order to be pleasant to the listener, the ending of a musical segment should resolve the musical tension created within the segment. Hence the last note of the musical segment tends to be tonally stable, i.e., a scale note, and in particular a tonic triad note. In addition, the beginning of a musical segment should provide the initial context for the listener to experience the music, and thus the first note also tends to be a tonally stable to match the final note. However, there is more flexibility (i.e., higher entropy) in the first note because the initial context could be provided in the first few notes (e.g., if the first note is an ornamental grace note).

This asymmetric information distribution within musical segments has important implications for research in music perception. For example, the asymmetric information distribution within words has been shown to be an important factor influencing how people process written words, such as the optimal viewing position (OVP) effect in reading isolated words, the preferred viewing location (PVL) effect in reading continuous texts, and visual field (VF) difference effects in word recognition (e.g., Brysbaert & Nozir, 2005; Brysbaert et al., 1996; Hsiao & Cheng, 2013; Legge et al., 1997). Similarly, the asymmetric information distribution within musical segments may also influence music notation reading behavior, such as eye movements in music reading (see, e.g., Madell & Hébert, 2008, for a review), and the asymmetric OVP pattern/visual field difference effect observed in reading music sequences (Wong & Hsiao, 2012). More specifically, Wong and Hsiao examined the OVP effect in reading three-note music sequences with a sequential matching paradigm and showed that music reading experts had the best performance when their initial fixation was directed to the sequence beginning, as compared to when it was to the middle or the ending of the sequence. In addition, participants had better performance when the music sequences were presented in the right VF (RVF) than in the left VF. Similar OVP and VF difference effects have also been reported in the recognition of English words, with the OVP located at the left of the word center (e.g., Brysbaert & Nazir, 2005; O'Regan, 1990; O'Regan et al., 1984), and a RVF advantage in English word recognition (e.g., Bradshaw & Gates, 1978; Brysbaert & d'Ydewalle, 1990; Brysbaert et al., 1996). Wong and Hsiao (2012) attributed the asymmetric OVP effect and RVF advantage in musical segment processing to the left-to-right reading direction in music reading, since most of the notes are typically recognized in the RVF with a left-to-right reading direction. Nevertheless, in visual word recognition, it has been shown that the asymmetric OVP effect and the RVF advantage can be accounted for by the information distribution within words (in addition to reading direction and hemispheric asymmetry; e.g., Brysbaert et al., 1996; Farid & Grainger, 1996; Hsiao & Cheng,

2013; see also Hsiao, 2011): participants typically have the best performance when the initial fixation is directed to the portion of the word with the most information, or a better performance when a word is presented in the visual hemifield in which the portion of the word with the highest information content is closer to the central fixation than when it is presented in the other visual hemifield. Thus, our current results suggest that the asymmetric OVP effect and the RVF advantage in musical segment processing observed in Wong and Hsiao (2012) may also be accounted for by the asymmetric information distribution within musical segments.

In addition to zeroth-order information distribution, here we also examined first-order information distribution within musical segments. We observed that in both information distributions, segment endings have the lowest entropy regardless of the segmentation method used. The low entropy at musical segment endings is consistent with the finding that, in Western music, melodies or musical phrases typically end with a stable harmonic tone, and thus phrase endings are typically more predictable (Aarden, 2003). This lower entropy/higher predictability at phrase endings than beginnings may also influence musical expectation in listeners. For example, Manzara et al. (1992) used the computer game *Chorale Casino* to investigate predictive probability of the melodies (derived from Bach Chorales 151 and 61) from human participants through a gambling game, and showed that phrase endings were typically associated with a higher predictability as compared with phrase beginnings or the middle of phrases. Witten et al. (1994) further showed that the entropy profile derived from human participants was very similar to that produced by a statistical model using 95 Bach chorale melodies (Conklin & Witten, 1995; see also Pearce & Wiggins, 2006), suggesting that musical expectation of humans is influenced by the statistics/regularities underlying the music the listeners are exposed to.

Note that these previous examinations of entropy profiles are calculated from the *conditional* probability of the predicted note given the observation of the previous notes; that is, the first-order (bigram) model (Conklin & Witten, 1995; Manzara et al., 1992; Pearce & Wiggins, 2006; Witten et al., 1994). The conditional probability is used because their aim is to investigate musical expectation while listening to a melody. In contrast, the entropy profiles in the zeroth-order model are calculated from the *marginal* note probabilities for each position of the segment. Hence, they measure the *a priori* information distribution within a musical segment before observing any notes, and thus are useful for studies of music perception behavior without contextual information, such as the OVP phenomenon observed in music reading (Wong & Hsiao, 2012). Here we show that in both cases, musical segment endings have the lowest information content regardless of the segmentation method used. The significant decrease in information content at musical segment endings may have been used by listeners implicitly for detecting segment boundaries.

We also examined the distributions of scale degrees within musical segments. We found that scale degrees 1 and 5 occur more frequently in the first and last note than in the middle of a segment. This is in contrast to other scale degrees, most of which occur more often in the middle of a segment than the beginning or the end (Figure 5). This difference in scale degree distribution may also be used by listeners as a cue to detect segment boundaries. Indeed, it has been shown that listeners use both pitch and temporal information in music phrase perception (Palmer & Krumhansl, 1987). Future work could further examine whether incorporating the information of scale degree distribution in music segmentation models (e.g., as in IDyOM) can better predict human music segmentation behavior.

In contrast to the entropy profiles of scale degrees, we found that the entropy profiles of pitch intervals were less consistent across song datasets and segmentation methods. Consistent with this finding, it has been shown that in processing unfamiliar melodies, people without music training tend to have difficulties in encoding pitch intervals as compared with pitch contours (e.g., Bartlett & Dowling, 1980; Cuddy & Cohen, 1976; Dowling, 1978; Fujioka et al., 2004). This result also suggests that pitch intervals may provide less useful information than contours for listeners to detect segment boundaries.

Finally, in the previous examinations of entropy profiles in melodies/music phrases (Conklin & Witten, 1995; Manzara et al., 1992; Pearce & Wiggins, 2006; Witten et al., 1994), the phrases used were typically longer than the musical segments we defined here. Multiple musical segments comprise a phrase, by analogy with multiple words comprising a phrase in language. Although musical segments are smaller units in music than phrases, the similarity in their entropy profiles, i.e., the decrease in entropy at the ending position, is intriguing and may suggest influence from similar factors, e.g., to have a closure with harmony/higher predictability to allow the listeners to respond. The cumulative moving average (CMA) entropy[5] for musical expectation obtained from

[5] The cumulative moving average entropy is the average of all the note entropies up until the current note position.

human data reported in these previous studies typically showed the highest entropy at the beginning with decreasing entropy towards the end. This is because the purpose of their experiments and models was to examine and simulate human musical expectation given a context: the beginning of a phrase typically had high entropy due to the lack of context; as more and more context revealed to support prediction, the predictability of notes gradually increased (i.e., the entropy of the conditional distribution gradually decreased). Nonetheless, the inverted U shape profile can be seen in the instantaneous entropy profiles of the musical segments[6] (as delineated by the fermatas, i.e., the prolonged notes) in the two Bach Chorales reported in Manzara et al. (1992), even as the CMA entropy decreases over time. According to our analysis reported here, the *a priori* entropy profile of musical segments without considering context is likely to have an inverted U shape. Hence, it would be intriguing to investigate the deviation of the entropy profile with context (i.e, the conditional distribution) from the *a priori* entropy profile without context as a measure of musical expectancy *within a musical segment* (cf. Abdallah & Plumbley, 2009; Dubnov, 2006, 2008).

In conclusion, we show that, similar to the information distribution within English words, musical segments also have an asymmetric information distribution, with higher entropy at sequence beginnings than sequence endings, although the asymmetry is not as dramatic as that within words. As the asymmetric information distribution within words has been shown to significantly influence how words are perceived and processed, this asymmetric information distribution within musical segments can also potentially modulate music reading behavior and thus should not be overlooked in the research on music perception.

Author Note

We are grateful to the Research Grant Council of Hong Kong (project # HKU 758412 H to J. H. Hsiao and project # CityU 123212 to A. B. Chan). The authors also thank the Wikifonia Foundation for providing the dataset, and Tuomas Eerola and Petri Toiviainen for the MIDI Toolbox. We thank the editor and three anonymous reviewers for helpful comments.

Correspondence concerning this article should be addressed to Janet Hsiao, Department of Psychology, University of Hong Kong, Pokfulam Road, Hong Kong. E-mail: abchan@cityu.edu.hk (A. B. Chan) or jhsiao@hku.hk (J. H. Hsiao)

References

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

Abdallah, S., & Plumbley, M. (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, *21*, 89-117.

Baayen, R. H., Pipenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bagnold, R. A. (1983). The nature and correlation of random distributions. *Proceedings of the Royal Society of London A*, *388*, 273-291.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 501-515.

Bourne, C. P., & Ford, D. F. (1961). A study of the statistics of letters in English words. *Information and Control*, *4(1)*, 48-67.

Bradshaw, J. L., & Gates, E. A. (1978). Visual field differences in verbal tasks: Effects of task familiarity and sex of subject. *Brain and Language*, *5*, 166-187.

Brent, M. R. (1999a). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, *34*(1), 71-105.

Brent, M. R. (1999b). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, *3*(8), 294-301.

Brysbaert, M., & d'Ydewalle, G. (1990). Tachistoscopic presentation of verbal stimuli for assessing cerebral dominance: Reliability data and some practical recommendation. *Neuropsychologia*, *28*, 443-455.

Brysbaert, M., & Nazir, T. (2005). Visual constraints in written word recognition: Evidence from the optimal viewing-position effect. *Journal of Research in Reading*, *28*, 216-228.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis - A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods*, *41*(4), 977-990.

Brysbaert, M., Vitu, F., & Schroyens, W. (1996). The right visual field advantage and the optimal viewing position effect: On the relation between foveal and parafoveal word recognition. *Neuropsychology, 10*, 385-395.

[6] In particular, the second sequence and thereafter. The beginning of the first sequence has high entropy due to the lack of context.

Carstairs-McCarthy, A. (2002). *An introduction to English morphology*. Edinburgh, UK: Edinburgh University Press.

Cambouropoulos, E. (1997). Musical rhythm: A formal model for determining local boundaries, accents and metre in a melodic surface. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 277-293). Berlin: Springer Verlag.

Cohen, P., Adams, N., & Heeringa, B. (2007). Voting experts: An unsupervised algorithm for segmenting sequences. *Intelligent Data Analysis*, *11*(6), 607-625.

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*, 51-73.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. Hoboken, NJ: Wiley-Interscience.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, *28*, 255-270.

de Nooijer, J., Wiering, F., Volk, A., & Tabachneck-Schijf, H. J. (2008). An experimental comparison of human and automatic music segmentation. In K. Miyazaki, M. Adachi, Y. Hiraga, Y. Nakajima, & M. Tsuzaki (Eds.), *Proceedings of the 10th International Conference on Music Perception and Cognition* (pp. 399-407). Sapporo, Japan: ICMPC.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341-354.

Dubnov, S. (2006). Spectral anticipations. *Computer Music Journal*, *30*(2), 62-83.

Dubnov, S. (2008). A unified view of prediction and repetition structure in audio signals with application to interest point detection. *IEEE Transactions on Audio, Speech and Language Processing*, *16*(2), 327-337.

Ducrot, S., & Pynte, J. (2002). What determines the eyes' landing position in words? *Perception and Psychophysics*, *64*, 1130-1144.

Eerola, T., & Toiviainen, P. (2004). *MIDI Toolbox: MATLAB tools for music research*. Jyväskylä, Finland: University of Jyväskylä: Kopijyvä. http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Elovitz, H., Johnson, R., McHugh, A., & Shore, J. (1976). Letter-to-sound rules for automatic translation of English text to phonetics. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *24*(6), 446-459.

Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language*, *53*, 351-368.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*(6), 1010-1021.

Hsiao, J. H. (2011). Visual field differences can emerge purely from perceptual learning: Evidence from modeling Chinese character pronunciation. *Brain and Language*, *119*(2), 89-98.

Hsiao, J. H., & Cheng, L. (2013). The modulation of stimulus structure on visual field asymmetry effects: the case of Chinese character recognition. *Quarterly Journal of Experimental Psychology*, *66*(9), 1739-1755.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, *27*(1), 75-97.

Krumhausl, C. K., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334-368.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An idealobserver model of reading. *Psychological Review*, *104*, 524-553.

Madell, J., & Hébert, S. (2008). Eye movements and music reading: Where do we look next? *Music Perception*, *26*, 157-170.

Manzara, L. C., Witten, I. H., & James, M. (1992). On the entropy of music: An experiment with Bach Chorale melodies. *Leonardo Music Journal*, *2*, 81-88.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, IL: University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. Chicago, IL: University of Chicago Press.

O'Regan, J. K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 395-453). New York: Elsevier Science.

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 250-257.

Palmer, C., & Krumhansl, C. L. (1987). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception and Psychophysics*, *41*, 505-518.

Pearce, M. T. (2014). *The IDyOM project*. London, UK: Queen Mary, University of London. https://code.soundsoftware.ac.uk/projects/idyom-project

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). Melodic grouping in music information retrieval: New methods and applications. In Z. W. Ras & A. Wieczorkowska (Eds.), *Advances in music information retrieval* (pp. 364-388). Berlin: Springer.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavior, and neural manifestations of musical expectation. *NeuroImage*, *50*, 302-313.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*, 377-405.

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, *30*(3), 277-292.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*, 21-30.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445-526.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, *4*, 525-553.

Rothschild, L. (1986). The distribution of English dictionary word lengths. *Journal of Statistical Planning and Inference*, *14*(2-3), 311-322.

Saffran, J. R., Newport, E. L., Aslin, R.N., Tunick, R. A., Barrueco, S. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606-621.

Schaffrath, H. (1995). The Essen folksong collection in kern format [Computer database, D. Huron, Ed.]. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379-423.

Shillcock, R., Ellison, T. M., & Monaghan, P. (2000). Eye-fixation behavior, lexical storage, and visual word recognition in a split processing model. *Psychological Review*, *107*, 824-851.

Shillcock, R. C., Hicks, J., Cairns, P., Chater, N., & Levy, J. P. (1996). Phonological reduction, assimilation, intra-word information structure, and the evolution of the lexicon of English: Why fast speech isn't confusing. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 233-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sleator, D., & Temperley, D. (2003). *The Melisma music analyzer*. Pittsburgh, PA: Carnegie Melon University. http://www.link.cs.cmu.edu/music-analysis/

Svartvik, J., & Quirk, R. (1980). A corpus of English conversation. Lund: LiberLaromedel Lund.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Temperley, D. (2014). Probabilistic models of melodic intervals. *Music Perception*, *32*, 85-99.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 394-402.

Trainor, L. J., & Trehub, S. E. (1993). Musical context effects in infants and adults: Key distance. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1-13.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Attention, Perception, and Psychophysics*, *56*(2), 125-132.

von Hippel, P. T. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception*, *17*, 315-327.

Witten, I. H., Manzara, L. C., & Conklin, D. (1994). Comparing human and computational models of music prediction. *Computer Music Journal*, *18*, 70-80.

Wong, Y. K., & Hsiao, J. H. (2012). Reading direction is sufficient to account for the optimal viewing position in reading: The case of music reading. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 2540-2545). Austin, TX: Cognitive Science Society.

Yannakoudakis, E. J., & Hutton, P. J. (1987). *Speech synthesis and recognition systems*. Chichester, UK: Ellis Horwood.

Yannakoudakis, E. J., & Hutton, P. J. (1992). An assessment of N-phoneme statistics in phoneme guessing algorithms which aim to incorporate phonotactic constraints. *Speech Communication*, *11*, 581-602.

Youngblood, J. E. (1958). Style as Information. *Journal of Music Theory*, *2*, 24-35.