Early Applications of Information Theory to Music

Marcus T. Pearce

Centre for Cognition, Computation and Culture, Goldsmiths College,
University of London, New Cross, London SE14 6NW
m.pearce@gold.ac.uk

March 28, 2007

The foundations of modern information theory were laid down by Hartley (1928) although it was to be twenty years before the first significant developments in the field were made with the publication of Claude Shannon's seminal paper on a mathematical theory of communication (Shannon, 1948). This work inspired a wave of interest throughout the 1950s in applying information-theoretic models to a number of fields ranging from psychology (e.g., Attneave, 1959) to computational linguistics (eg., Shannon, 1951). It is interesting to note that the new methods were applied to music as early as 1955 (see Cohen, 1962). Of particular relevance to music scholars was that portion of the theory that pertains to discrete noiseless systems and, in particular, the representation of such systems as stochastic Markov sources, the use of *n*-grams to estimate the statistical structure of the source and the development of *entropy* as a quantitative measure of the uncertainty of the source. Inspired perhaps by the use of entropy to estimate the fundamental uncertainty of printed English (Shannon, 1948, 1951), researchers used information-theoretic concepts and methods throughout the 1950s and 60s both to analyse music (Cohen, 1962; Meyer, 1957) and to generate new compositions (e.g., Brooks Jr., Hopkins, Neumann & Wright, 1957; Hiller & Isaacson, 1959; Pinkerton, 1956). In this review, we focus on the use of information-theoretic methods in quantitative analyses of music referring the reader to existing reviews of synthetic and compositional applications (Ames, 1987, 1989; Cohen, 1962; Hiller, 1970).

In one of the first such studies, Pinkerton (1956) computed a monogram distribution of diatonic scale degrees in a corpus of 39 monodic nursery rhymes yielding a redundancy estimate of 9%. Following a similar approach, Youngblood (1958) examined the entropy of two different musical styles: first, 20 songs in a major key from the Romantic period (composed by Schubert, Mendelssohn and Schumann); and second, a corpus of Gregorian chant. Zeroth- and first-order distributions of chromatic scale degrees were computed from these corpora; the latter exhibited much higher redundancy than the former indicating that the pitch of a note is highly constrained by knowing the pitch of the previous note. Furthermore, while redundancy differed little between the three Romantic composers, the overall redundancy of this corpus was lower than that of the Gregorian chant.

More detailed information-theoretic studies of musical style were conducted under the supervision of Lejaren Hiller at the University of Illinois. Hiller & Bean (1966), for example, examined four sonatas composed by Mozart, Beethoven, Berg and Hindemith respectively. Each sonata was segmented analytically and monogram distributions of chromatic pitch classes were computed for each segment. The results indicated that average entropy increases (and redundancy decreases) from the Mozart to the Beethoven example, from the Beethoven to the Hindemith example and from the Hindemith to the Berg. Other stylistic differences emerged from more detailed comparisons of the entropy and redundancy figures for individual segments. Hiller & Fuller (1967) extended this approach in an analysis of Webern's symphony (Op. 21) in two directions: first, they computed first- and second- as well as zeroth-order entropy estimates (notes occurring simultaneously were flattened in order of pitch height); and second, they examined intervallic and rhythmic representations as well as pitch. The symphony was divided into three sections (exposition, development and recapitulation) each of which was examined separately. The authors were able to relate differences in entropy and redundancy between the three sections to differences in structural complexity of the musical features examined. However, the study also highlighted the effects of sample size on the reliability of estimated probabilities as well as the effects of alphabet size on the generality of the estimates.

These early studies may be criticised on a number of grounds, the first of which relates to the manner in which probabilities are estimated from the samples of music (Cohen, 1962). It is generally assumed that a distribution estimated from a sample of music constitutes an accurate reflection of a listener's perception of the sample. However a listener's perception (e.g., of the first note in the sample) cannot be influenced by music she has not yet heard (e.g., the last note in the sample) and her state of knowledge and expectation will change dynamically as each note in the music is experienced (Meyer, 1957). In order to address concerns such as these, Coons & Kraehenbuehl developed a system of calculating dynamic measures of information (predictive failure) in a sequence (Coons & Kraehenbuehl, 1958; Kraehenbuehl & Coons, 1959). However, it remains unclear whether the method could be computationally implemented and its application generalised beyond the simple examples given. Furthermore, like the studies reviewed above, the method fails to reflect the fact that a listener hears a piece of music in the context of extensive experience of listening to other pieces of music (Cohen, 1962).

A second criticism of these early studies is that they are generally limited to low fixed-order estimates of probability and therefore do not take full advantage of the statistical structure of music. A final criticism relates to the representation of music (Cohen, 1962). With the exception of Hiller & Fuller (1967), all of these studies focused exclusively on simple representations of pitch ignoring other features or dimensions of the musical surface and interactions between these dimensions. Even Hiller & Fuller (1967) had to consider each dimension separately since they had no way of combining information derived from different features.

The use of information-theoretic concepts and methods in psychology lost favour during the so-called "cognitive revolution" of the late 1950s and early 1960s that saw the end of behaviourism and the birth of artificial intelligence and cognitive science (Miller, 2003). This loss of favour was based partly on objective inadequacies of Markov chains as models of psychological representations and of language in particular (Chomsky, 1957). However, it seems likely that it was also due, in part, to an arbitrary association of information-theoretic analysis with behaviourism and the fact that corpus size and the complexity of statistical analyses were necessarily limited by the processing power of the computers available. Nonetheless, the knowledge engineering approach to examining mental representations and processes became the dominant paradigm in cognitive science until the 1980s when a resurgence of interest in connectionist modelling (Rumelhart & McClelland, 1986) stimulated a renewed emphasis on learning and the statistical structure of the environment.

These trends in cognitive-scientific research had a knock-on effect in music research. Connectionist models of musical structure and music perception began to be examined in the late 1980s (Bharucha, 1987; Desain & Honing, 1989; Todd, 1988). However, with a handful of isolated exceptions (e.g., Baffioni, Guerra & Lalli, 1984; Coffman, 1992; Knopoff & Hutchinson, 1981, 1983; Snyder, 1990), it was not until the mid 1990s that information-theoretic methods and statistical analyses again began to be applied to music (Conklin & Witten, 1995; Dubnov, Assayag & El-Yaniv, 1998; Hall & Smith, 1996; Ponsford, Wiggins & Mellish, 1999; Reis, 1999; Triviño-Rodriguez & Morales-Bueno, 2001). Instrumental in this regard was the fact that many of the limitations of the early efforts were addressed by Darrell Conklin's development of sophisticated statistical models of musical structure (Conklin, 1990; Conklin & Cleary, 1988; Conklin & Witten, 1995).

In particular, the predictive systems developed by Conklin consist of a long-term component that is derived from a large corpus of music and a short-term component that is constructed dynamically for each musical work: the estimated probability of a given event at a given point in the work reflects the combined action of these two models. Furthermore, each model uses n-grams of a number of different orders up to a global bound in computing its probability estimates. In more recent work, the maximum order is allowed to vary depending on the context (Pearce & Wiggins, 2004). Finally, the system can compute distinct probability distributions for different features or dimensions of the musical surface, weight them according to their relative entropy and combine them in arriving at a final probability estimate in a given context. Various kinds of interaction between different features can be explicitly represented and exploited in estimating probabilities.

References

Ames, C. (1987). Automated composition in retrospect: 1956–1986. *Leonardo*, *20*(2), 169–185.

Ames, C. (1989). The Markov process as a compositional model: A survey and tutorial. *Leonardo*, *22*(2), 175–187.

Attneave, F. (1959). *Applications of information theory to psychology*. New York: Holt.

Baffioni, C., Guerra, F., & Lalli, L. (1984). The theory of stochastic processes and dynamical systems as a basis for models of musical structures. In M. Baroni & L. Callegari (Eds.), *Musical Grammars and Computer Analysis* (pp. 317–324). Leo S. Olschki.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*(1), 1–30.

Brooks Jr., F. P., Hopkins, A. L., Neumann, P. G., & Wright, W. V. (1957). An experiment in musical composition. *IRE Transactions on Electronic Computers*, *EC-6*(1), 175–182.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Coffman, D. D. (1992). Measuring musical originality using information theory. *Psychology of Music*, *20*, 154–161.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, *7*(2), 137–163.

Conklin, D. (1990). *Prediction and Entropy of Music*. Master's dissertation, Department of Computer Science, University of Calgary, Canada.

Conklin, D. & Cleary, J. G. (1988). Modelling and generating music using multiple viewpoints. In *Proceedings of the First Workshop on AI and Music* (pp. 125–137). Menlo Park, CA: AAAI Press.

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Coons, E. & Kraehenbuehl, D. (1958). Information as a measure of structure in music. *Journal of Music Theory*, *2*(2), 127–161.

Desain, P. & Honing, H. (1989). The quantisation of musical time: A connectionist approach. *Computer Music Journal*, *13*(3), 56–66. Also in Music and Connectionism, 1991, P.M. Todd and D.G. Loy (Eds., MIT Press.

Dubnov, S., Assayag, G., & El-Yaniv, R. (1998). Universal classification applied to musical sequences. In *Proceedings of the 1998 International Computer Music Conference* (pp. 332–340). San Francisco: ICMA.

Hall, M. & Smith, L. (1996). A computer model of blues music and its evaluation. *Journal of the Acoustical Society of America*, *100*(2), 1163–1167.

Hartley, R. V. L. (1928). Transmission of information. *Bell System Technical Journal*, *7*, 535–563.

Hiller, L. (1970). Music composed with computers – a historical survey. In H. B. Lincoln (Ed.), *The Computer and Music* (pp. 42–96). Cornell, USA: Cornell University Press.

Hiller, L. & Bean, C. (1966). Information theory analyses of four sonata expositions. *Journal of Music Theory*, *10*(1), 96–137.

Hiller, L. & Fuller, R. (1967). Structure and information in Webern's Symphonie, Op. 21. *Journal of Music Theory*, *11*(1), 60–115.

Hiller, L. & Isaacson, L. (1959). *Experimental Music*. New York: McGraw–Hill.

Knopoff, L. & Hutchinson, W. (1981). Information theory for musical continua. *Journal of Music Theory*, *25*, 17–44.

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

Kraehenbuehl, D. & Coons, E. (1959). Information as a measure of the experience of music. *Journal of Aesthetics and Art Criticism*, *17*(4), 510–522.

Meyer, L. B. (1957). Meaning in music and information theory. *Journal of Aesthetics and Art Criticism*, *15*(4), 412–424.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.

Pearce, M. T. & Wiggins, G. A. (2004). Rethinking Gestalt influences on melodic expectancy. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the Eighth International Conference of Music Perception and Cognition* (pp. 367–371). Adelaide, Australia: Causal Productions.

Pinkerton, R. C. (1956). Information theory and melody. *Scientific American*, *194*(2), 77–86.

Ponsford, D., Wiggins, G. A., & Mellish, C. (1999). Statistical learning of harmonic movement. *Journal of New Music Research*, *28*(2), 150–177.

Reis, B. Y. (1999). *Simulating Music Learning with Autonomous Listening Agents: Entropy, Ambiguity and Context*. Doctoral dissertation, Computer Laboratory, University of Cambridge, UK.

Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. Cambridge, MA: MIT Press. Volumes 1 and 2.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423 and 623–656.

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell System Technical Journal*, *30*, 50–64.

Snyder, J. L. (1990). Entropy as a measure of music style: The influence of a priori assumptions. *Music Theory Spectrum*, *12*(1), 121–160.

Todd, P. M. (1988). A sequential neural network design for musical applications. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School* (pp. 76–84). San Mateo, CA: Morgan Kaufmann.

Triviño-Rodriguez, J. L. & Morales-Bueno, R. (2001). Using multi-attribute prediction suffix graphs to predict and generate music. *Computer Music Journal*, *25*(3), 62–79.

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