Statistical Explorations of Harmony, Cadences, and Form in the Annotated Mozart Dataset

The corpus-based analysis of musical structures (including harmony, melody, counterpoint, and rhythm) has gained increasing importance over the past decade. Apart from studies of tonal harmony in general, a growing number of theoretical, computational, and cognitive studies examine the use of cadences in the "classical" repertoire (e.g., Bigo et al., 2018; Duane & Jakubowski, 2018; and Sears et al., 2018). However, computational research on musical structures is still hindered due to (1) the paucity of available datasets and (2) the lack of shared standards for encoding music analyses (e.g., Rizzo & Marsden, 2018).

In an attempt to address these lacunae, this paper examines cadences, their harmonic characteristics, and their form-functional correlations (Hepokoski & Darcy, 2006) in a new dataset consisting of all keyboard sonatas by Wolfgang Amadé Mozart. This dataset contains expert annotations at three analytic levels:

First, comprehensive harmonic analyses drawing on a novel annotation standard (a regular expression formalizing a modified Roman numeral system) that provide information on key, chordal root, chord form, chord position, and local voice-leading (Neuwirth et al., 2018); second, annotations of more than 1.000 cadence tokens based on a cadence typology developed in various music theoretical studies (Caplin, 1998; Neuwirth & Bergé, 2015; and Sears, 2018); and third, analyses of musical form based on recent music theoretical work (Caplin, 1998; Hepokoski & Darcy, 2006; Diergarten & Neuwirth, 2018) and represented by a new annotation standard combining elements from MEI and TEI.

Given that algorithmic cadence finding is still a challenging and by no means flawless task (e.g., Bigo et al., 2018), we propose a semi-automated approach, one that combines a labeled dataset (cadence type and endpoint) with an implemented heuristics that helps delineate cadence instances both from one another and from non-cadential harmonic progressions. This allows us to characterize the frequency of cadence types, their harmonic design, their probabilistic profiles, and their form-functional uses.

(1) The paper shows that cadence types differ significantly with regard to their frequency distribution. This concerns especially the two types of authentic cadences (perfect and imperfect), which are often described as roughly equal alternatives in harmony textbooks that would differ only with respect to the final scale degree in the soprano. Contrary to this view, our corpus study demonstrates that the perfect authentic cadence is by far the most frequent cadence type (accounting for 46% of all cadence tokens). In comparison, the imperfect authentic cadence is the least important type, appearing only 48 times (5%) and resembling the set of failed cadences (13%) in terms of frequency and formal function. The second most frequent type is the half cadence, covering 36% of the dataset.

(2) Our study further shows that the cadence types differ not only with respect to their frequency of appearance, but also – and contrary to textbook accounts – with regard to their harmonic layout. For instance, imperfect authentic cadences

generally do not make use of extended predominant chords, cadential 6/4 chords, and penultimate dominant chords featuring a dissonant seventh.

(3) Using Bayesian statistics, we examine the hypothesis that certain chords (especially the ii^6 chord) function as reliable predictors of complete cadential progressions; this hypothesis posited for "classical" music in general (Caplin, 1998) can be confirmed with regard to Mozart's keyboard sonatas.

(4) Our analysis shows that the choice of specific subtypes of half cadences and failed cadences is not random in Mozart's music but is correlated with the formal function at hand. For instance, deceptive cadences more likely occur within a sonata form's primary-theme zone than in the secondary-theme section; the reverse is true of evaded cadences (Caplin, 1998; Hepokoski & Darcy, 2006). Further, the number of cadences increases in the secondary-theme section in comparison to earlier formal zones; the number of cadences decreases significantly in development sections. These findings largely conform to music theoretical hypotheses.

To summarize, this paper aims to provide an empirically testable basis for the use of cadences in the classical style, using Mozart's keyboard sonatas as a point of departure on which future studies can build. More generally, the symbolic dataset presented in this paper can be used to train, evaluate, and improve computational models for examining musicological research questions.

References

Bigo, L., et al. (2018). "Relevance of Musical Features for Cadence Detection." In *Proceedings of the 19th International Society for Music Information Retrieval Conference* (ISMIR), Paris 2018. Ed. E. Benetos, E. Gómez, X. Hu, and E. Humphrey.

Caplin, W. E. (1998). *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press.

Diergarten, F. and Neuwirth, M. (2018). *Formenlehre. Ein Lese- und Arbeitsbuch zur Instrumentalmusik des 18. und 19. Jahrhunderts* (= Grundlagen der Musik, vol. 7). Laaber: Laaber.

Duane, B. and Jakubowski, J. (2018). "Harmonic Clusters and Tonal Cadences: Bayesian Learning Without Chord Identification." *Journal of New Music Research* 47/2, 143–165.

Hepokoski, J. and Darcy, W. (2006). *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. New York: Oxford University Press.

Neuwirth, M. and Bergé, P. (2015). *What is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*. Leuven: Leuven University Press.

Neuwirth, M., et al. (2018). "The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets." *Frontiers in Digital Humanities* 5:16.

Rizo, D. and Marsden, A. (2018). "An MEI-based Standard Encoding for Hierarchical Music Analyses." *International Journal on Digital Libraries* (https://doi.org/10.1007/s00799-018-0262-x).

Sears, D., et al. (2018). "Simulating Melodic and Harmonic Expectations for Tonal Cadences Using Probabilistic Models." *Journal of New Music Research* 47/1, 29–52.