Chapter 5

Expertise in Action: LTWM as Control Over Simulation

In this chapter, I examine one aspect of music theory expertise: the formation and manipulation of a schema interpretation. Here, I argue that forming and changing an interpretation is a type of memory expertise which represents control over simulation in WM—essentially, the selection and instantiation of different property and relation simulators over time from LTM into WM. In the first part of the chapter, I discuss issues in interpretation formation, particularly the contribution of top-down and bottom-up factors to ambiguous figure perception. I then present a case study examining perceptual ambiguity of the Modulating Prinner Schema in Mozart piano sonatas. Here, I discuss Mozart's use of the Modulating Prinner and Step-Descent Fauxbourdon Romanesca schemata within sonata form transitions, which I argue presents a case for perceptual ambiguity between these two schemata. I conclude by discussing the results from a qualitative survey on ambiguous figure perception in the transition of Mozart's Piano Sonata no. 2, iii. The results demonstrate that the transition is perceptually bistable, but that the extent to which an interpretation can be formed or changed depends on the level of expertise with Galant schemata categories.

Issues in Top-Down Effects in Interpretation and Perception: Hazy Boundaries Between Perception and Cognition

One of the primary issues for debate in cognitive sciences is the notion of cognitive penetrability—the idea that higher level cognitive processes can *directly* affect lower-level perceptual processing. Such higher-level processes include things like beliefs, desires, emotions, intentions, and linguistic representations (Firestone and Scholl 2016, 1). In other words, those in

favor of the cognitive penetrability hypothesis argue that higher-level states directly affect what we see at the perceptual level, which calls for a radical non-modular blurring of perception and cognition. Research examining cognitive penetrability has primarily focused on the visual domain, with authors pointing to phenomena such as ambiguous figure perception as evidence for top-down effects on interpretation formation in perception (see Long and Toppino 2004 for a review). While the consensus is generally that ambiguous figure interpretation involves a mixture of top-down and bottom-up processing (e.g., Intaite *et al.* 2013; Leptourgos *et al.* 2020), recent research has strongly refuted the cognitive penetrability hypothesis, claiming that there exists little to no evidence for top-down effects on perception (see Firestone and Scholl 2014; 2015a,b,c; 2016). Instead, these authors suggest effects found in previous research do not represent so-called top-down effects on perception, but rather demonstrate confined effects of cognition—the formation of a judgment in higher-level processing as opposed to a direct modulation of perceptual processes (Firestone and Scholl 2016, 9).[1] Effects on interpretation are driven primarily by higher-level cognitive processes involving attention and memory (*Ibid.* 11, 15), which Firestone and Scholl argue are self-contained modular cognitive processes.

Regardless of one's position on cognitive penetrability, the scholarship in this area does support the notion that interpretation of a percept is heavily dependent on effects of attention and memory (see Toppino 2003; Meng and Tong 2004; Pearson and Brascamp 2008). Changing interpretation in bistable visual stimuli is mediated by a shifting of visual attention and fixation to different parts of the stimulus (see Taddei-Ferretti *et al.* 2008) and is also affected by effects

[1] Essentially, the notion here is that the formation of an interpretation depends primarily on cognitive processes, and that these higher-level processes do not affect lower-level perceptual ones (i.e., the way in which the eye and lower-level visual system, for example, takes in and processes information is unaffected by cognition. This information remains the same, and what changes are higher-level cognitive processes). Some authors argue that this definition of perception is far too narrow (see Cañal-Bruland, van der Kamp and Gray 2016).

of memory priming (Hortliz and O'Leary 1993). Other research suggests an important role for imagery in ambiguous figure perception, such that those with higher imagery ability demonstrate greater ease in generating, maintaining, and alternating interpretations (Pearson, Clifford and Tong 2008). For the current project, the primary concern entails effects of attention and memory in percept interpretation. Here, I suggest that forming and changing an interpretation are inherently part of music theory memory expertise, which is used in both imagery and perception. Forming an interpretation in this context is therefore understood to operate through top-down control and selection of representations present in simulator pools—essentially, the selective instantiation of different property and relation simulators from LTM into WM. The extent to which this process is largely under *active* control of the central executive or governed by more automated 'bottom-up' perceptual priming is contingent on the type of activity (i.e., imagery versus perception).

Control over Interpretation in Perception: Musical Ambiguous Figures and the Case of the Modulating Prinner

Controlling an interpretation in perception, as compared to imagery, presents a multitude of challenges. Unlike in imagery, where the object for interpretation can be more easily held and manipulated in WM, forming an interpretation—let alone changing one—in auditory perception requires much more automation because WM is constantly being filled with new sensory information. Forming or modulating an interpretation in auditory perception therefore relies much more extensively on attention: shifting attention can be one means of selecting a different simulator base from LTM. This does not rule out an important role for imagery during perceptual interpretation, however. As one example, I will show that theorists can make use of simultaneous imagery in the form of subvocalization (musical and/or verbal) to bring different simulator pools

online. Combined with the ability to further focus attentional resources via simulation using music theoretic concepts, theorists are able to toggle between interpretations and evaluate them during listening. Here, I illustrate this phenomenon by reporting the results from a qualitative survey on musical ambiguous figure perception. In the first section, I discuss the musical materials and contexts—modulating Prinners and Step-Descent Romanescas in sonata form transitions—used in the study. I then outline the study's guiding questions, hypotheses, method, and results, followed by a discussion that contextualizes my findings within the framework developed in this dissertation.

Modulating Prinner and Sonata Form: An Overview

The modulating Prinner is a subtype of the general schema, which is heard as ending in the key of the dominant. It was a commonly used and very useful schema serving to facilitate a modulation (Gjerdingen 2007, 52). It therefore is an ideal schema to use in formal locations where modulation is expected, such as transitional spaces of sonata forms. The sonata form transition (TR) is a space located between the primary (P) and secondary (S) theme areas in a two-part exposition. It culminates in a punctuation: a medial caesura, more often than not one that is modulatory, before the entrance of S in a new key. In major mode sonatas, the first-level default for modulatory transitions is a V: HC MC, and in minor mode sonatas it is either a III: HC or v: HC MC (Hepokoski and Darcy 2006, 25-26). Should the transition be non-modulatory, the second-level default for medial caesuras is a I: HC MC (*ibid.,* 26). The rarest third- and fourth- level defaults for medial caesuras are the V: PAC in modulatory transitions and I: PAC MC in non-modulatory transitions, respectively (*ibid.*, 27, 29).

Hepokoski and Darcy (2006) discuss the typical structure of the TR space as one that presents many analytical challenges due to the ambiguous nature of the P-TR boundaries,

particularly in the case of fused or merged transitions (Hepokoski and Darcy 2006, 95).[2] In general, however, transitional rhetoric is fairly distinct from the rhetoric of the P space, including normative expectations for energy gain through to the arrival of the medial caesura. Thus, while the opening of the TR space may be inherently ambiguous, transitional rhetoric is more easily identifiable as one gets closer to the medial caesura. Typical textural features of transitional rhetoric include *Fortspinnung*, sequential activity, *forte* dynamics, and drive toward a structural dominant. The drive toward the medial caesura has several recognized stages. Firstly, the push toward arrival of a structural dominant, which is often prolonged through a dominant pedal (Hepokoski and Darcy 2006, 30-31). This is followed by a continuation of energy gain through increased dynamics, rhythmic activity, and the like (*ibid.*, 33). Subsequently, the arrival of the medial caesura often occurs with several *forte* hammer blows to mark the accumulation of energy gain toward its terminal peak (p. 34), followed by a grand pause, sometimes including *caesura* fills. This is followed by the launching of S, the secondary theme area, which is exemplified by a sudden change in texture and, often, *piano* dynamics in the second expositional key, which together confirm the previous material as the medial caesura (p. 36). The TR space in the recapitulation is a passage that invites re-composition. As the S reappears in the tonic key, a modulatory transition is not required. However, composers often play with modulatory expectations, including 'superfluous' or 'unnecessary' re-composition in the P-TR spaces. This area is often subjected to reinterpretation involving tonal shifts, often in the subdominant direction, before 'correcting' the modulation back to the home key (*Ibid.,* 236). Such procedures, both in the expositional and recapitulatory TR spaces, provide numerous interpretational

[2] Hepokoski and Darcy note that composers likely understood the TR space as one that was appended to the P-space as continuation modules (p. 93).

challenges, including ‘perceptually ambiguous’ areas where multiple interpretational options exist, in both formal and tonal interpretation.

The rhetorical structure of the TR space is one that encourages particular pairings of schemata over time. In modulatory transitions, this often includes the pairing of the modulating Prinner, to initiate the tonal shift towards the dominant (in major-mode sonatas), with a converging cadence and Ponte to punctuate and extend the dominant arrival before the medial caesura. Such common pairings in the TR space are reflected in the transitional probabilities between the Prinner, modulating Prinner, and other schemata typically used in transitions as Gjerdingen (2007, Chapter 27) discusses (see Figure 5.1). The Prinner and Modulating Prinner schemata are most likely to be followed by cadences, half-cadences, Ponte, Cudworth, and converging cadence schemata, demonstrating the flexibility of Prinner schema in terms of tonal goals as it may shift toward closure either in the key of the tonic or in the dominant. Modulating, dominant trajectory Prinner pairings (e.g., Modulating Prinner to Converging) are particularly probabilistic in transitional spaces, which has been proposed (Byros 2011) as one that facilitates 18th century interpretational responses, such as the perception of tonal modulation *prior* to receiving any new key information. In such instances, the modulating Prinner-converging cadence pair is contained within a large-scale Indugio schema, initiating perception of modulation, and shifting expectations towards a dominant arrival in the key of the dominant before the modulation is marked by a change of key (see Figure 5.2). Thus, while the introduction of key-confirming pitch information occurs in the final two stages of the modulating Prinner, the 18th century interpretation is to perceive the change of key at the opening of the modulating Prinner, snapping the perception of tonal orientation to a new one when perception of the schema first occurs.

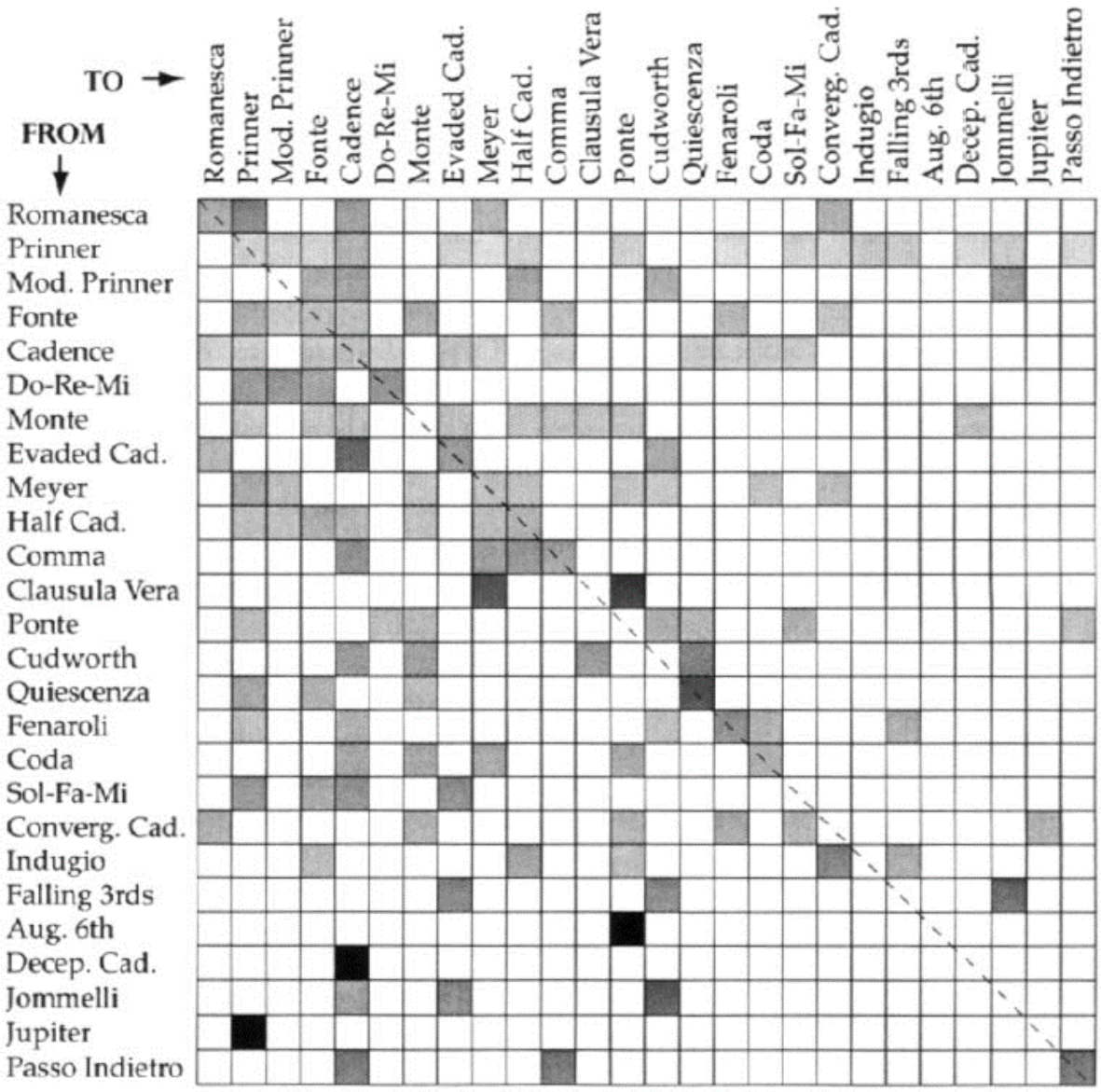


FIGURE 5.1: Fig 27.1 from Gerjdingen 2007, p 372

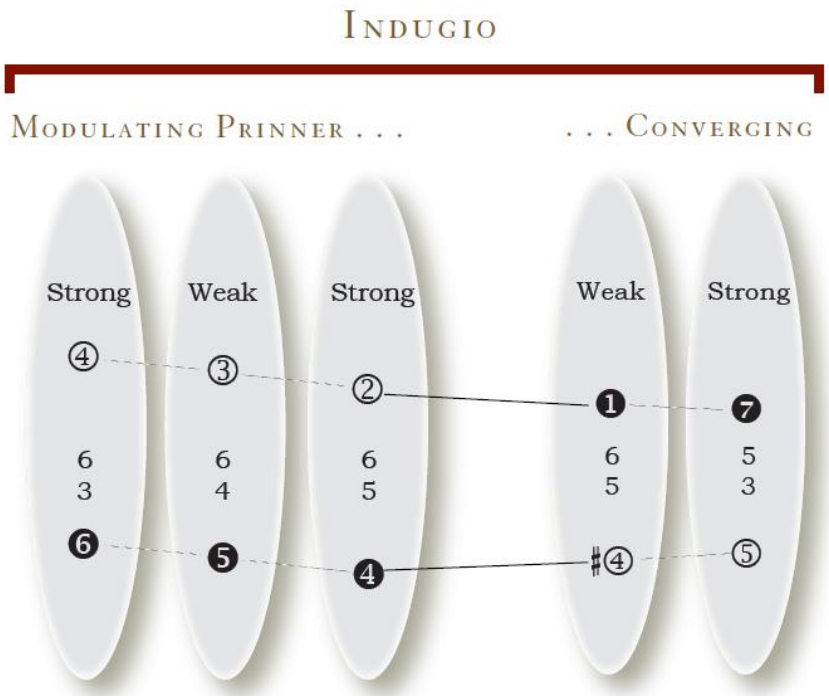


FIGURE 5.2: From Byros (2011)

A Mozart Case Study: Two Transitional Schema Options

The transitional space in a sonata form offers an abundance of interpretational affordances, particularly around the medial caesura goal. As each sonata essentially contains two versions of the transition—one in the exposition which may or may not modulate, and one often re-composed version in the recapitulation that does not modulate—several mental 'templates' exist for transitions in these contexts, both as a generalized 'transition-concept' (i.e., particular simulators activating over time), and on a piece-by-piece basis. In the exposition, just as there are multiple options for tonal goal of the transition, there are also, therefore, multiple potential 'pathways' through tonal space which drive toward the medial caesura. As composers often play with these expectations, the sonata form transition is one in which there are multiple 'schema-templates' available to the analyst, affording some flexibility in interpretation in these spaces. In

this section, I will discuss two different schema options for modulating and nonmodulating transitions—the Modulating Prinner and the Step-Descent Fauxbourdon Romanesca (hereby Step-Descent Romanesca). I will begin by discussing the role of the Modulating Prinner in modulating transitions, followed by the Step-Descent Romanesca in tonic-confirming (nonmodulating) contexts.

Modulating Prinner

The modulating Prinner, as previously discussed, is one of the primary schemata seen when modulating to the dominant, and therefore often occurs in modulating transitions whose goal is the key of the dominant (most typically V: HC MC). The modulating Prinner, in addition to the tonal pliability in scale degree identity, has several other features as discussed by Gjerdingen (2007). Given the position of the modulating Prinner in the transition, the structure is often one of sequential or *Fortspinnung*-like construction. In addition, a typical feature of the modulating Prinner is an octave leap in the upper voice, which functions as a signal to change scale degree identity (where $\hat{3}$ becomes $\hat{6}$ in the new key, Gjerdingen 2007, 355). Here I will discuss two versions of such a modulating Prinner in Mozart piano sonatas K. 280, iii and K. 309, i.

Piano Sonata no. 2, K. 280, iii. The third movement of K. 280 is a *Presto*, Type 3 sonata, featuring an independent, modulating transition ending with a V: HC medial caesura. The boundary between the P and TR space is slightly obscured: the P space ends with a I: PAC in measure 16, and the TR begins the following measure taking the form of a dissolving P-codetta type over a tonic pedal point (see Figure 5.3). Rhetorically, the TR space is clearly initiated by measure 25, which aligns with the presence of the modulating Prinner, complete with *forte* dynamics, octave-leaps in the upper voice, and sequential, fragmentary construction typified by a continuation phrase. It is here at measure 25 that the modulating Prinner marks the availability of the new key of the dominant (C major), which is reinforced by measure 27 with the occurrence of the new leading tone, and fully confirmed with the presence of the dominant arrival in C major in measure 32. As discussed in Byros (2011), the ordering of Galant schemata is typical of Mozart: a modulating Prinner (measures 25-28) leads to a converging cadence (measures 31-32), which is prolonged until the medial caesura proper (measure 37). Together, both schemata form a larger scale Indugio, which prolongs $\hat{4}$ (measures 25-29), passing through the converging cadence to reach the dominant of the key. The medial caesura here is marked by a grand pause, which then continues to the S-theme in the key of the dominant (C major), starting as is typical on *piano* dynamic.

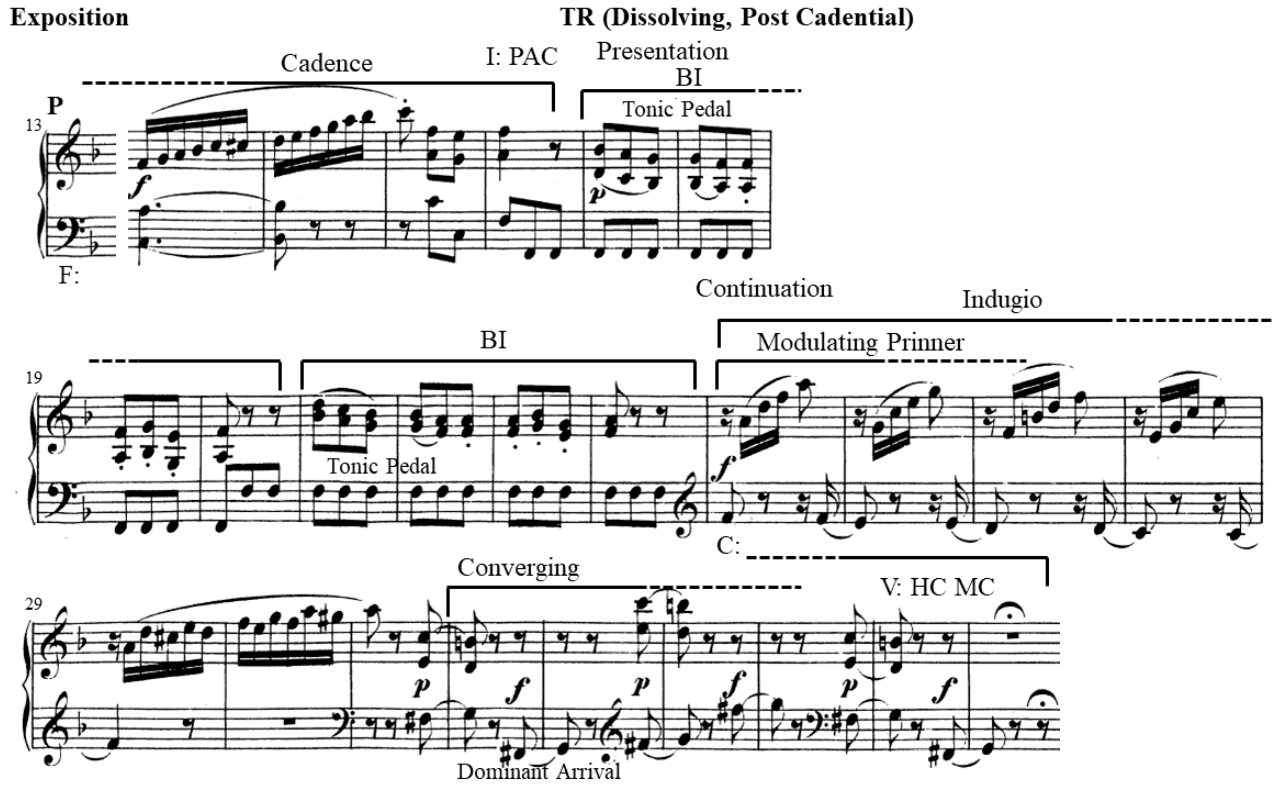


FIGURE 5.3 The recapitulatory transition (measures 132-148) receives only a small adjustment to re-orient the listener and accommodate the lack of modulation in preparation for the S-theme. Here, the original modulating Prinner occurs in measure 132 (C major), which is then followed by a secondary modulating Prinner in the home key, which re-orients the listener back to F major (see Figure 5.3). The crux is at measure 140, which continues as the expositional TR: a large scale Indugio contains the second modulating Prinner (measures 136-139) and converging cadence (measures 142-148).

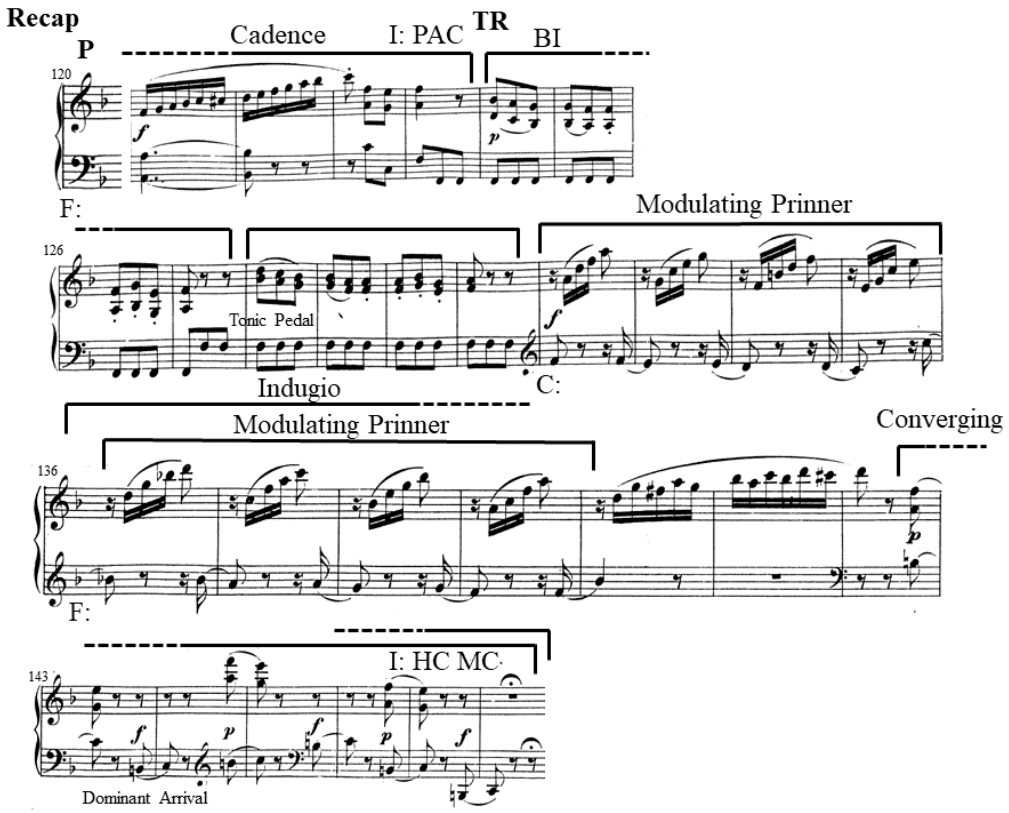


FIGURE 5.4: Recapitulatory Transition of K. 280, iii, m 120-148.

Piano Sonata no. 10, K. 309, i. Similar to K. 280, the first movement of K. 309 features an independent, modulating transition, facilitated by a modulating Prinner which leads to a V: HC MC (see Figure 5.5). In this movement, however, the boundary between the P and TR spaces is more clearly articulated, featuring an elided I: PAC in measure 21 marking the end of the P and beginning of TR. This boundary is also marked by typical TR rhetoric, including *forte* dynamics and sequential construction. And much as was seen in K. 280, this marks the beginning of the modulating Prinner which facilitates the modulation to the key of the dominant. This modulating Prinner also contains a similar octave leap construction, with a two-bar sequential stage presentation. The parallel tenth voice structure is emphasized at the beginning of each stage; however, the bassline is modified to reflect the mixing of the Prinner and Fenaroli schemata, which further emphasizes the tonic arrival in the new key in measure 27. This arrival is slightly undermined by the repeated iteration of $\hat{4}$ (from the previous stage) above the final stage, only reaching resolution in measure 29. This is followed by a two-bar modified Pulcinella schema (measures 29-30), and the V: HC MC in measure 32. Mozart includes *caesura* fill in measures 33-34, at the beginning of the S in measure 35 in the key of the dominant, marked with *piano* dynamics.

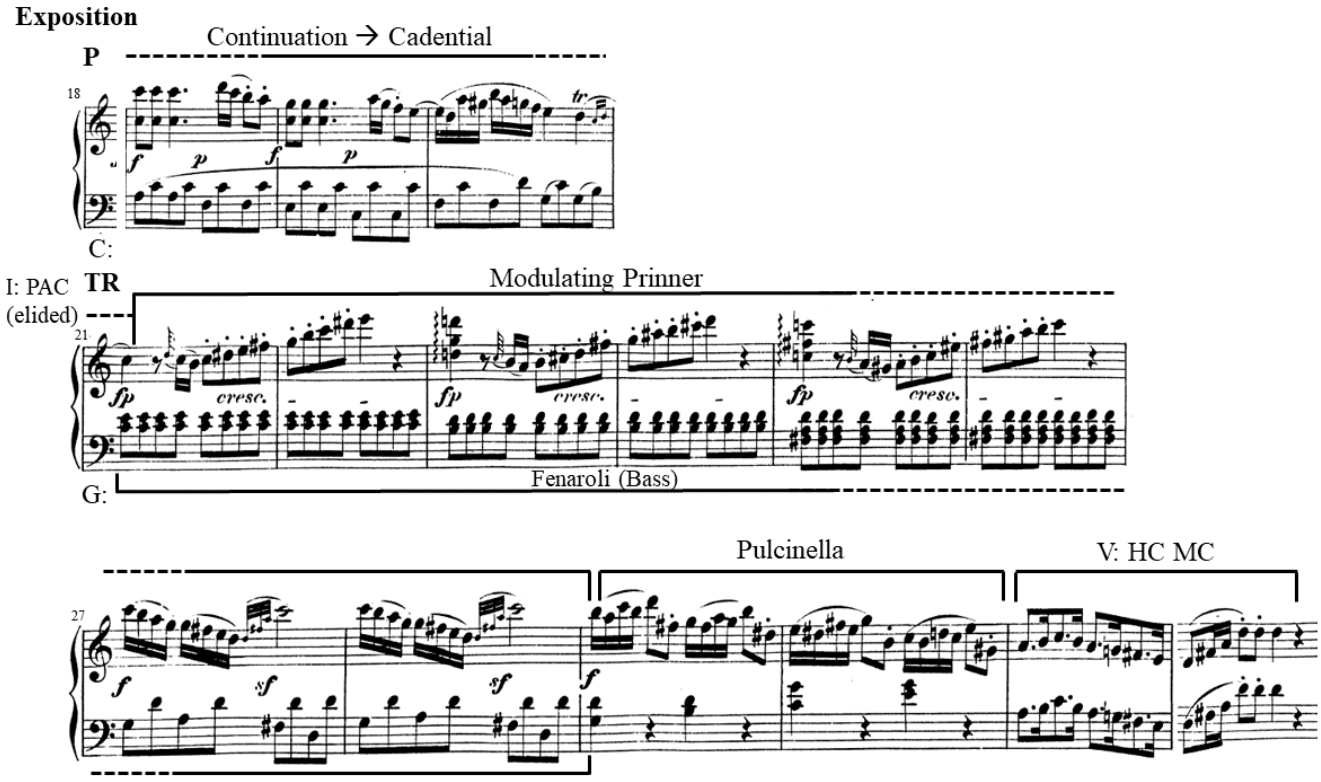


FIGURE 5.5: Expositional Transition from K. 309, i, m 18-32.

The recapitulatory TR is quite similar to that found in the exposition, with a slight modification to adjust for the lack of modulation (see Figure 5.6). The schematic layout is like the exposition, with the same Modulating Prinner and Fenaroli bassline in measures 116-124. The final stage of the Prinner includes a pedal point rather than tonic-dominant alternations and includes the melodic portion of the Pulcinella schema from the exposition, essentially merging these two areas into one. Within the context of the modulating Prinner version heard in the exposition, and the continued presence of F-sharp leading tones, the pedal is not yet reinterpreted as a dominant arrival, even though rhetorically the presence of a pedal point in this section of the TR space is highly indicative of dominant arrival. Instead, the crux occurs at measure 125, which acts as a perceptual snap back to a dominant interpretation to align with the structure of the

expositional TR. The I: HC MC is maintained here in measures 125-126, followed by caesura fill and the tonic key entrance of S in measure 129.

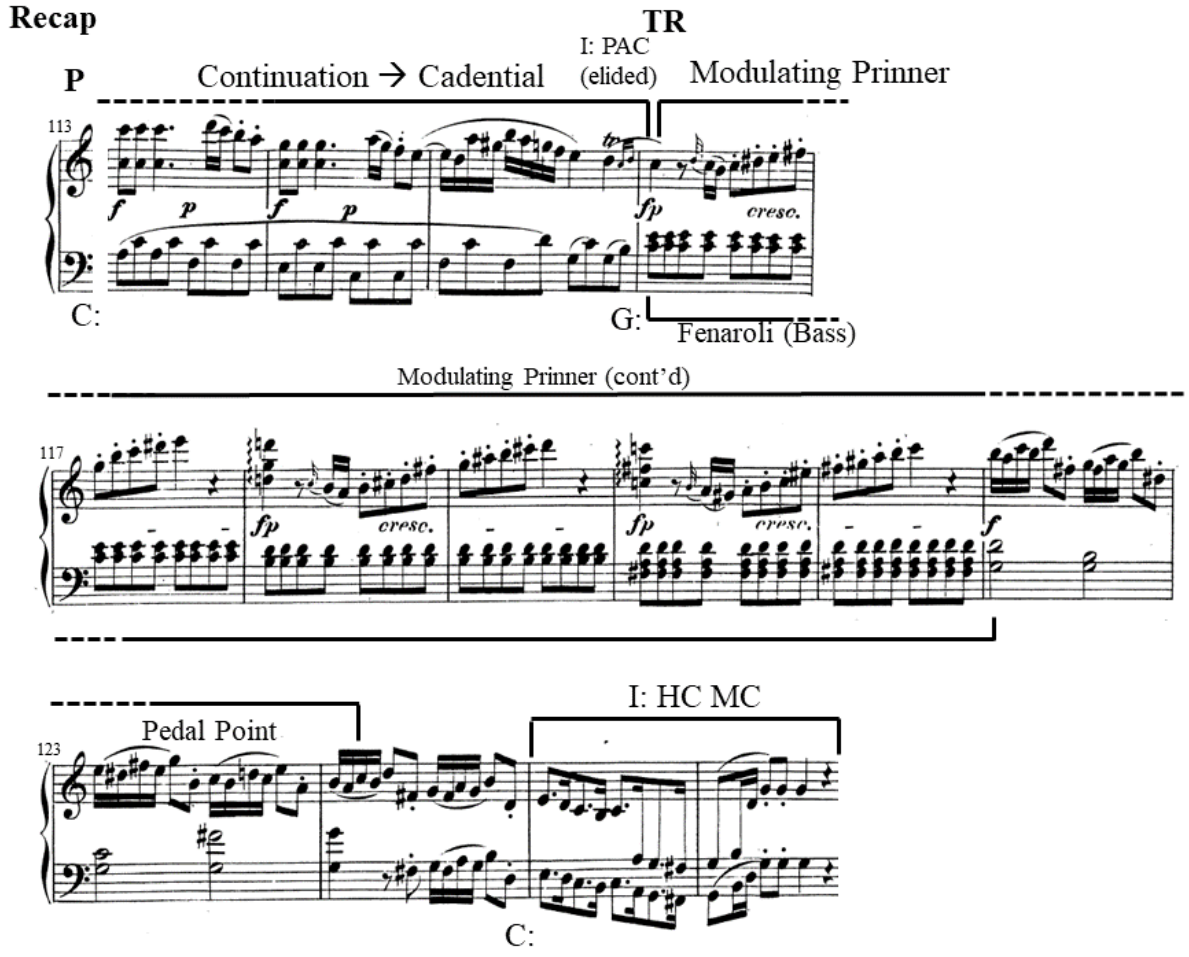


FIGURE 5.6: Recapitulatory Transition of K. 309, i, m 113-126.

Step-Descent Fauxbourdon Romanesca

The Romanesca schema is typically used as an opening gambit and often occurs as a basic idea in thematic material. There are several variants of the Romanesca. The earliest form is the leaping, descending 3rds sequential variant (Gjerdingen 2007, 29), and the second, a scalar step-descent variant which descends the entire octave from tonic to tonic, alternating 5/3 and 6/3 harmonies (*Ibid.*, 32). Both of these variants' sequential structure afford functional usage in

continuation phrases. The last variant of the Romanesca, a Galant hybrid, combines the leaping and stepwise variants, and forms the version commonly used as an opening gambit (*Ibid.,* 33). There is some structural similarity between the step-descent Romanesca and the Prinner, including the stepwise bassline and alternating 5/3 and 6/3 harmonies. The Prinner's distinguishing feature, however, is the parallel-tenth counterpoint in the outer voices; the Romanesca's upper voice motion, by way of contrast, typically alternates between tonic and dominant scale degrees. This can be more closely approximated in the step-descent Romanesca by the addition of fauxbourdon voice leading and harmonic inversion, similar to voicings used in the Rule of the Octave (Gjerdingen 2007, 468). This structural similarity is evident in the use of the Step-Descent Fauxbourdon Romanesca as a key-confirming schema in both non-modulating and modulating transitions. Compared to the modulating Prinner, the Step-Descent Romanesca is used to reinforce an ongoing key, rather than initiate a modulation. Below I present two versions of such a schema in the TR space, to demonstrate its typical usage in such spaces.

Piano Sonata no. 15, K. 576, i. The first movement of Mozart's Piano sonata K. 576 features a non-modulating, independently thematized transition. Here, a Step-Descent Romanesca is used in the transition in a similar manner to the modulating Prinner (in ordering Romanesca-Converging-Ponte), on the way toward the medial caesura (Figure 5.7). The transition begins in measure 16—and much like the first movement of K. 309—features an elided PAC following the end of the P. The first eight measures of the transition firmly iterate the home key of D major with two repetitions of the Aprile (measures 16-19) and Fenaroli (measures 20-23) schemata. This is followed by the Step-Descent Romanesca (measures 24-25), descending nearly the entire octave in the bassline (from $\hat{3}$ to $\hat{4}$), and the entire octave in the soprano ($\hat{1}$ to $\hat{1}$). Contextually, it shares many similarities to the Modulating Prinner, including the placement and progression of the schema to a converging cadence and Ponte, driving toward the I: HC MC. But it also shares several rhythmic and motivic similarities to the Modulating Prinner previously discussed, such as the octave displacement in the soprano voice, and rhythmic figuration similar to that used in K. 280.

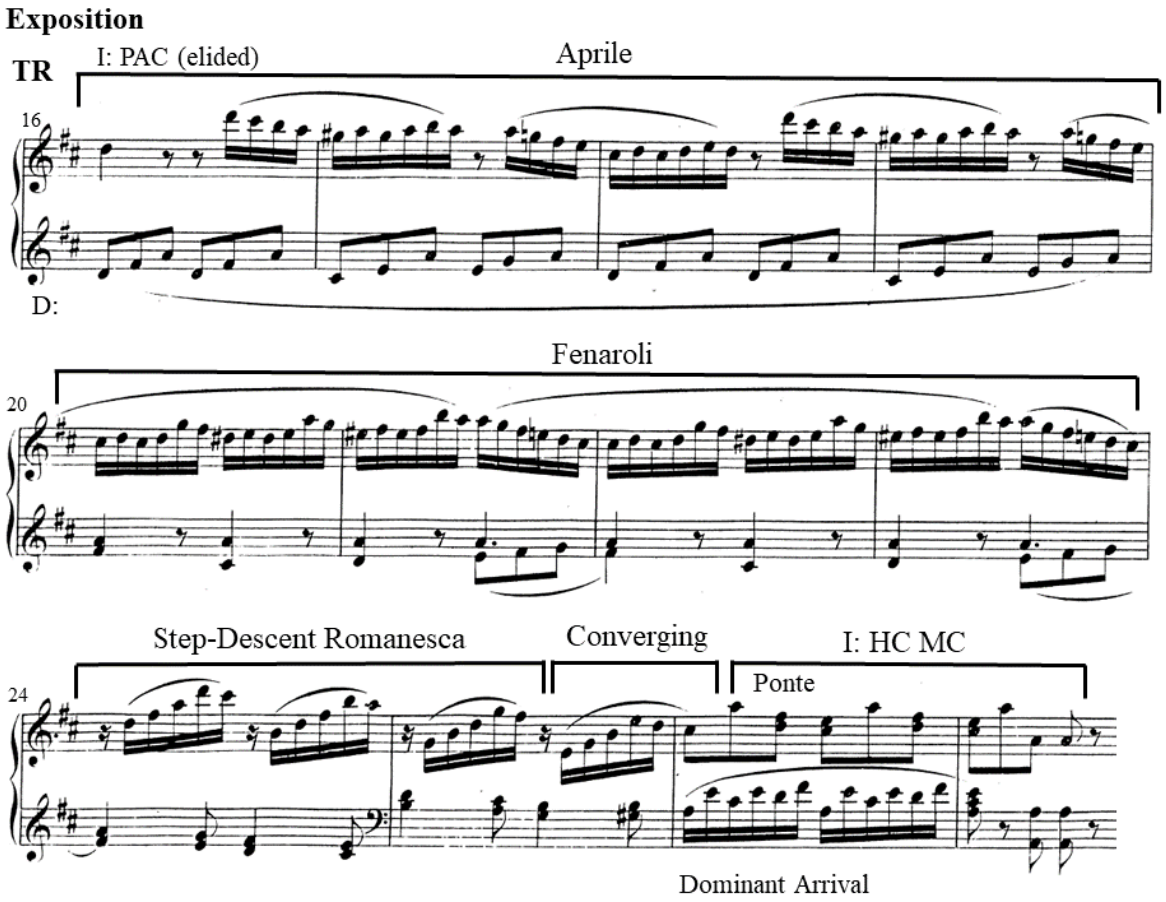


FIGURE 5.7: Expositional Transition of K. 576, i, m 16-27.

The recapitulatory TR is recomposed to resemble a dissolving P-type, which articulates several key areas, including the subdominant, before returning to the home key for a I: HC MC (see Figure 5.8). The TR begins in measure 106, at the point in the exposition where a restatement of P was provided (measure 9). Here, the P-material shifts to a subdominant inflection for the contrasting-idea in measures 110-112, which is followed by fragmentation and imitative material in the keys of the minor dominant (a) and submediant (b). The crux occurs at measure 118 where the original Step-Descent Romanesca is provided, which functions to suddenly snap the back to the home key of D major and is followed by the same Converging Cadence to Ponte I: HC MC in measures 119-121.

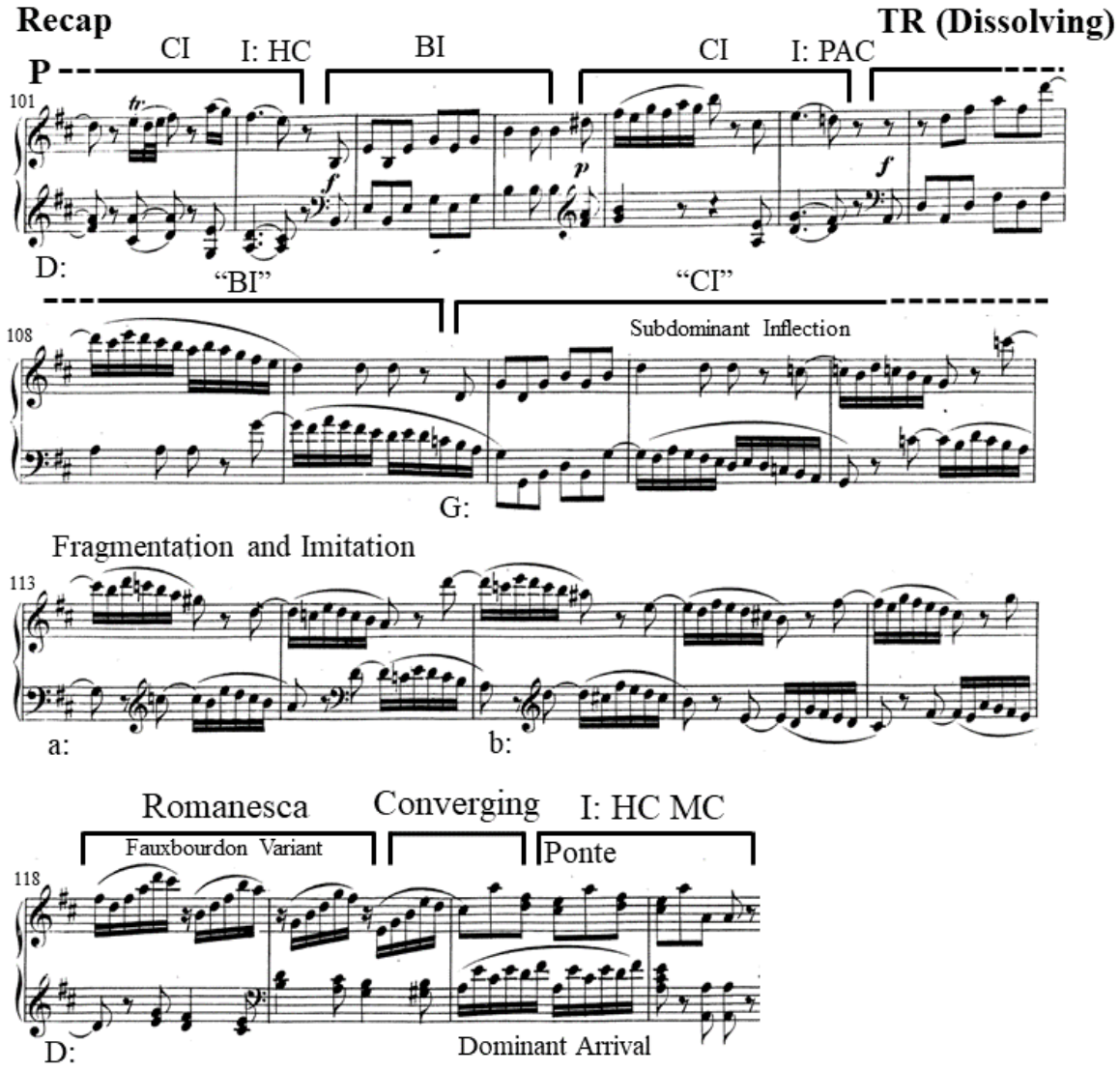


FIGURE 5.8: Recapitulatory Transition of K. 576, i, m 101-121.

Piano Sonata no. 14, K. 310, iii. The third movement of K. 310 presents an interesting case. The final movement is a Type 4 sonata, or sonata rondo, which uses a trimodular block (TMB) strategy in the exposition, and a medial caesura declined in the recapitulation. Mozart uses a Step-Descent Romanesca as tonic confirmation within the TMB, as well as a hybrid modulating Prinner at the second medial caesura in the exposition (see Figure 5.9).

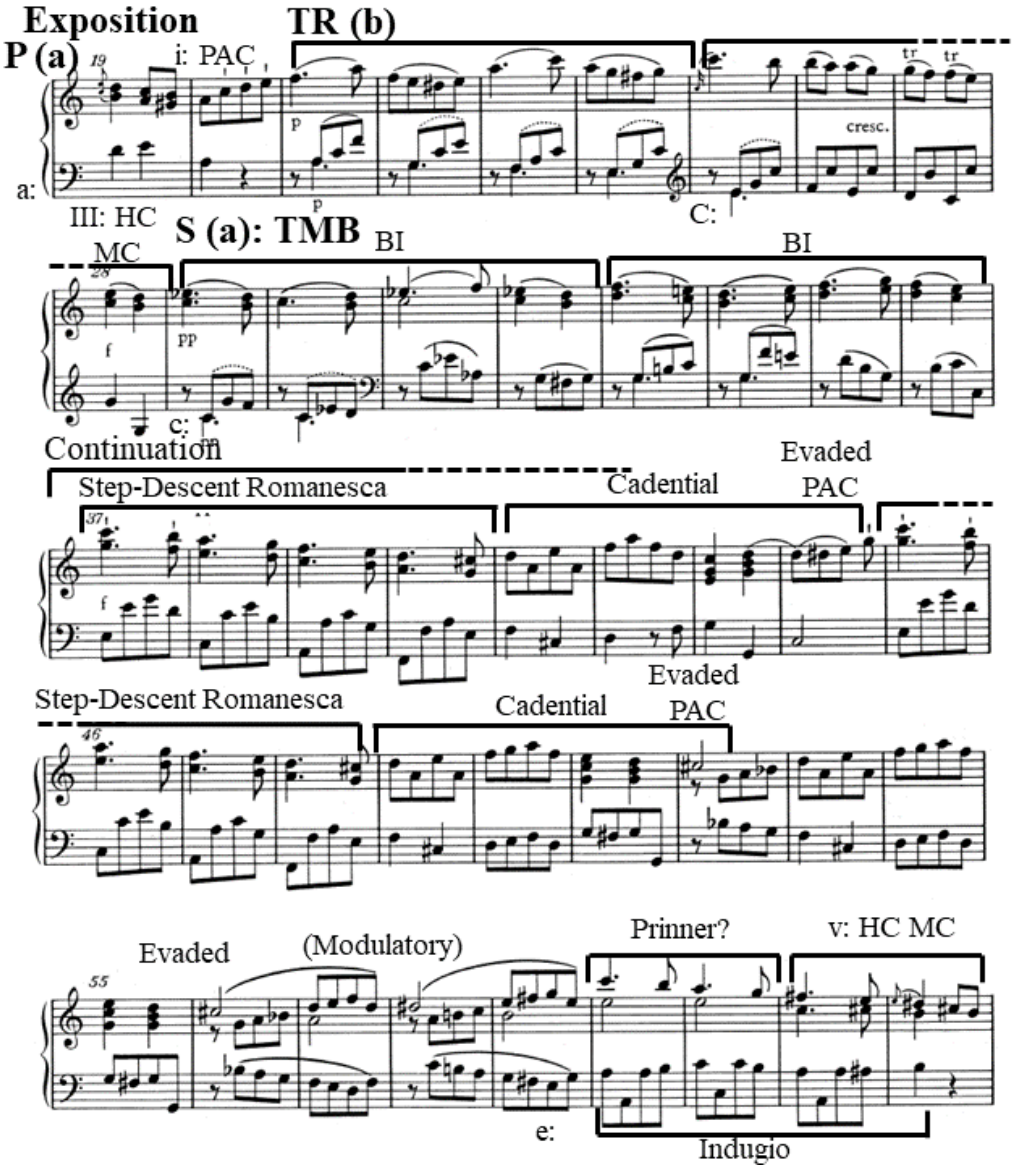


FIGURE 5.9: Expositional Trimodular Block (TMB) in K. 310, iii, m 19-63.

Here, the modulating transition (or b rotation) begins after a i:PAC, which quickly modulates to the major mode mediant, and weakly articulates a III: HC MC in measure 28. The S zone, or third rotation (a), begins directly afterwards (in the minor mediant), and is structured as a sentence, in which the Step-Descent Romanesca functions as the continuation phrase, in the process confirming the key of C major. This schema is repeated twice, each time leading to an evaded cadence, which is followed by modulatory material leading to a second medial caesura in the key of the minor dominant. Interestingly, this second articulation of the MC is structured as an Indugio in which the upper voice (which had previously been the descending line, starting on tonic, in the Step-Descent Romanesca) is reinterpreted in the new key as a minor mode Prinner soprano line (le-sol-fa-me) in measures 60-61.

This procedure is like other usages by Mozart, in which the modulating Prinner and Converging cadences are embedded into an Indudio (see K. 280 above). This sort of compressed, fused hybrid is also typical of Mozart, where the Converging cadence (bass line) is combined with the melody of the modulating Prinner (Gerjdingen 2007, 353). This recasting of the primary voice from the non-modulating Step-Descent Romanesca to a modulating Prinner variant highlights the plasticity and combinatorial potential of the different parts of these schemata. The recapitulatory version of the transition includes a medial caesura declined, without the expositional TMB (see Figure 5.10). Here, the modulating hybrid Indugio (Modulating Prinner + Converging) is excluded, but the tonic confirming Step-Descent Romanesca is maintained.

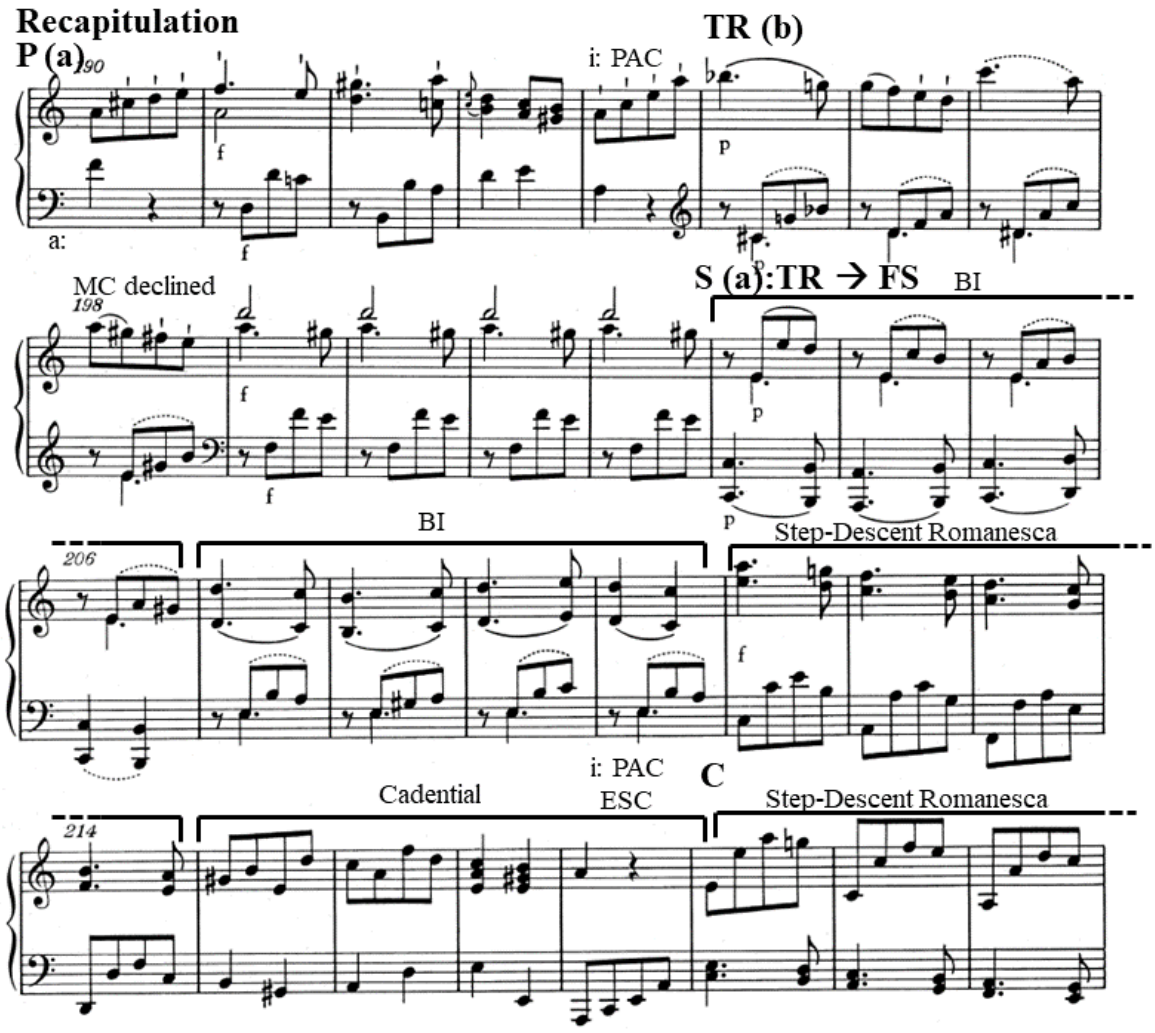


FIGURE 5.10: Recapitulatory Transition of K. 310, iii, m 190-221.

An Experimental Verification: Effects of Attention, Memory, and Expertise on Bistable Perception in Mozart's Piano Sonata no. 2, K. 280, iii

I have demonstrated that both the modulating Prinner and Step-Descent Romanesca occur within TR spaces in several of Mozart's piano sonatas. These schemata have several features in common, which suggests potential for perceptual ambiguity:

- Sequential organization and presentation of schemata stages, which features fragmentary-like construction typical of medial (continuation) function.

- Several melodic and rhythmic similarities: a rhythmically offset melodic voice that arpeggiates against a bass that is metrically stable.
- Parallel tenths between two of the three or four voices present.
- Similar spatial-temporal locations (i.e., in the TR space leading up to the medial caesura), and temporal orderings with other schemata (e.g., _____—Converging—Ponte).

However, important differences help distinguish one schema from the other. For example, the distinguishing Modulating Prinner features are:

- Equal-length schema stages that range from one to two measures long. This type of construction is ideal for providing the listener with time to perform a tonal reorientation as each stage is presented for enough time such that scale degrees and harmonies can be reinterpreted in the new key.
- Stages that are harmonized with a mixture of 5/3 and 6/3 sonorities which alternate.
- Temporal-spatial location near the beginning of the transition as it often initiates transitional rhetoric.

Whereas contrastingly, the distinguishing features of the Step-Descent Romanesca are:

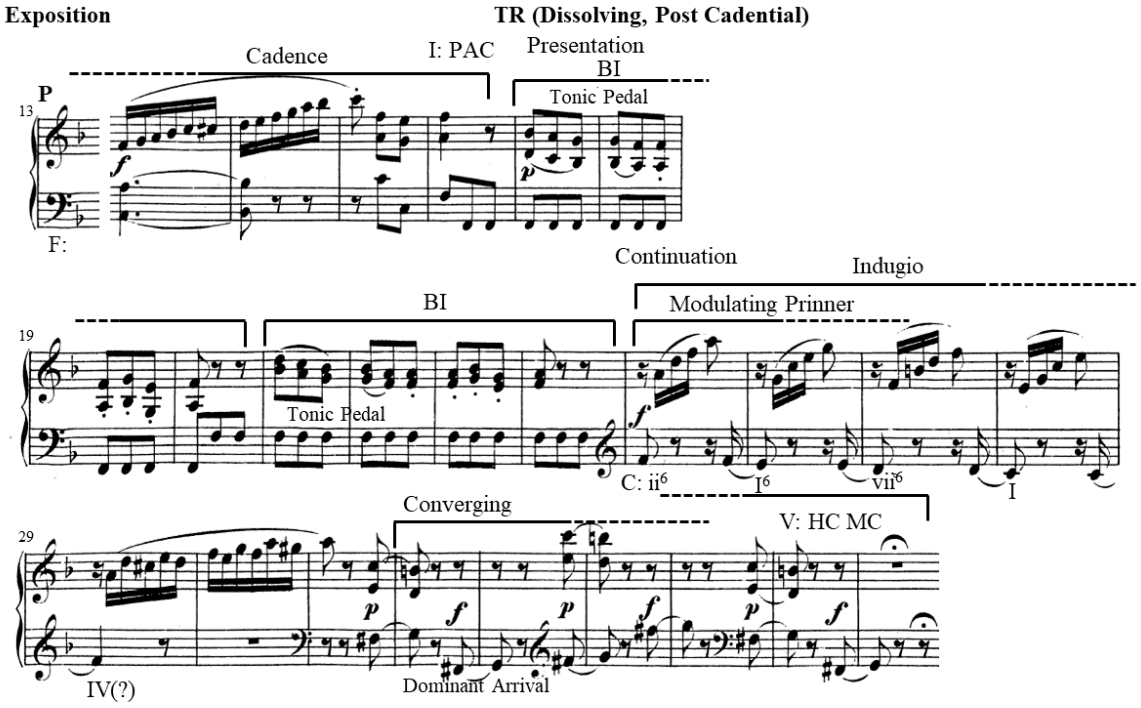
- Twice the number of schema stages as the Prinner (eight in total), descending the entire scale from tonic to tonic.
- Shorter and rhythmically unequal stages (e.g., long-short-long-short). This feature aligns with the tonic-confirming function of the Step-Descent Fauxbourdon Romanesca, with which there is no need for the composer to offer time to the listener to tonally reorient.
- Only 6/3 harmonies, does not alternating with 5/3 harmonies.
- Occurs near the middle or end of the transition, and it therefore does not initiate transitional rhetoric.

Given the overlapping features of these two schemata, there is a case to be made for perceptual ambiguity on the basis of 'bottom-up' features alone. However, an interpretation may also vary with expertise. Byros (2009a,b; 2012) has argued that expertise with Galant schemata categories provides a means to access eighteenth century modes of hearing—particularly as it relates to tonality perception. An examination of reception history of Beethoven's *Eroica* reveals three primary strains of tonal interpretation: one which perceives a modulation to G-minor, a "Cloud" strain which proposes a form of tonal ambiguity, and lastly one which reads the whole passage within the key of E-flat major. Byros demonstrates that the tendency to perceive key change in these opening bars operates on a historical axis, indicating that the historical situation or context is a contributing factor in the variation of key perception. The G-minor hearing therefore appears to a historically and stylistically appropriate one, one which Byros proposes arises from the perception of the *le-sol-fi-sol* schema.

Expertise with Galant schemata therefore appears to provide listeners with the ability to rapidly reorient tonally and perceive 'unconfirmed' modulations, which from a modern more monotonal perspective, are likely perceived as temporary tonicizations that merely function to prolong an ongoing key, rather than alter it. This suggests that modern listeners—lacking in familiarity with Galant schemata—are more likely to perceive tonal reorientation in modulatory passages as occurring when new key information is introduced (e.g., accidentals) in the signal ('bottom-up'). Contrastingly, those with schemata expertise are more likely to associate particular schema—understood as co-occurrences of scale degrees and specific harmonies—as belonging to a particular key, even when there have yet to be any changes in accidentals. Therefore, there are two possible interpretations of the transition from the Presto of K. 280: an historically informed one, where an early modulation is perceived as occurring with the onset of

a modulating Prinner (see Figure 5.11a,b), and a more modern one where the modulation is not perceived until the introduction of new key information (specifically, $\sharp\hat{4}$ in the new key) (see Figure 5.12a,b). However, there exists a grey area in this second interpretation in light of the overlapping features between the Prinner and Step-Descent Romanesca schemata discussed above. The question, therefore, is: is it possible, with at least some expertise in Galant schemata, to hear both Modulating Prinner and Step-Descent Romanesca interpretations in this transition?

(a). Expositional Modulating Prinner



(b). Recapitulatory Modulating Prinners

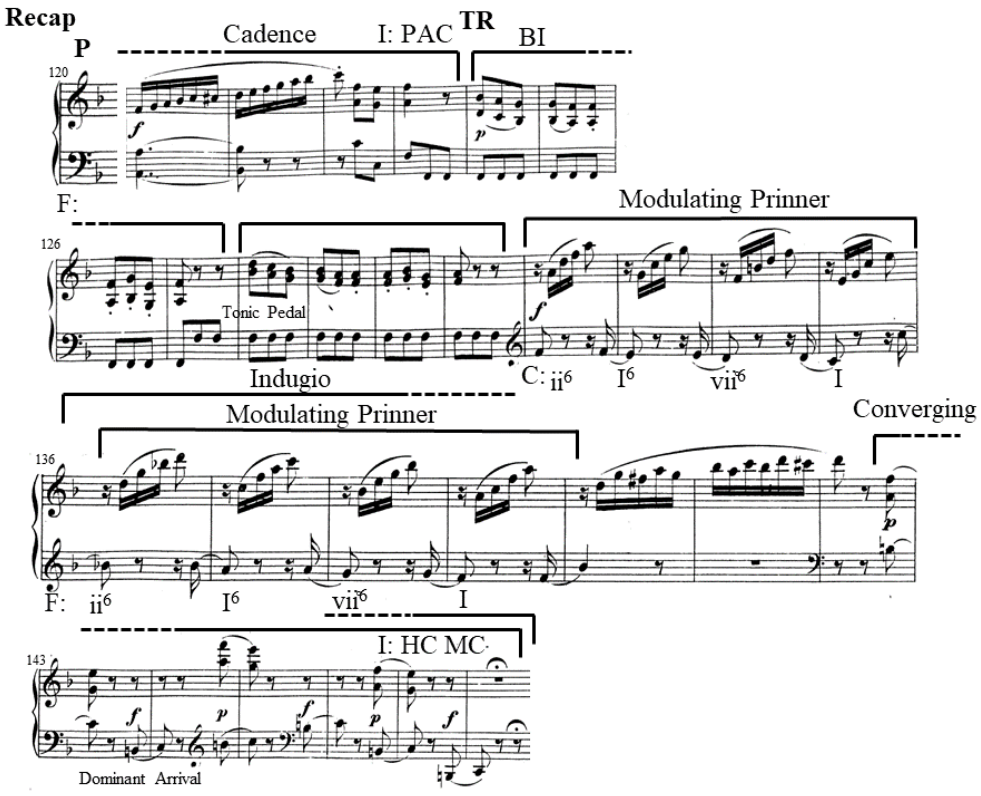
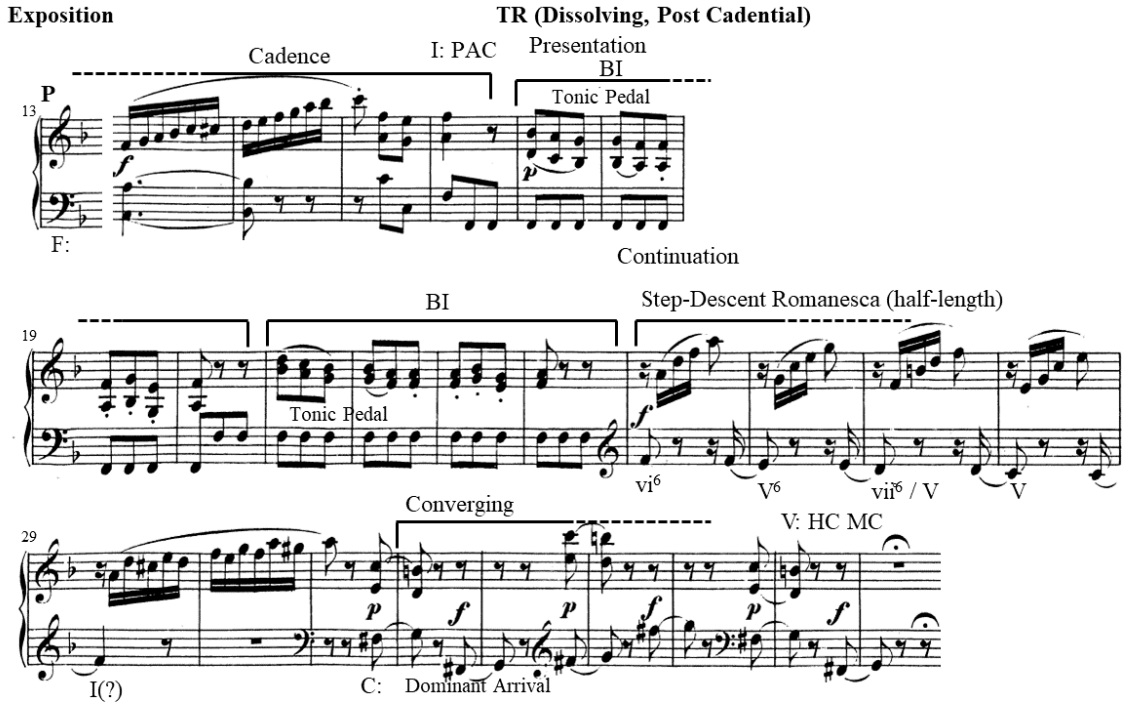


FIGURE 5.11a,b: Modulating Prinner Interpretations in the Exposition (a) and Recapitulation (b) of K. 280, iii

(a). Expositional Step-Descent Romanesca



(b). Recapitulatory Step-Descent Romanesca

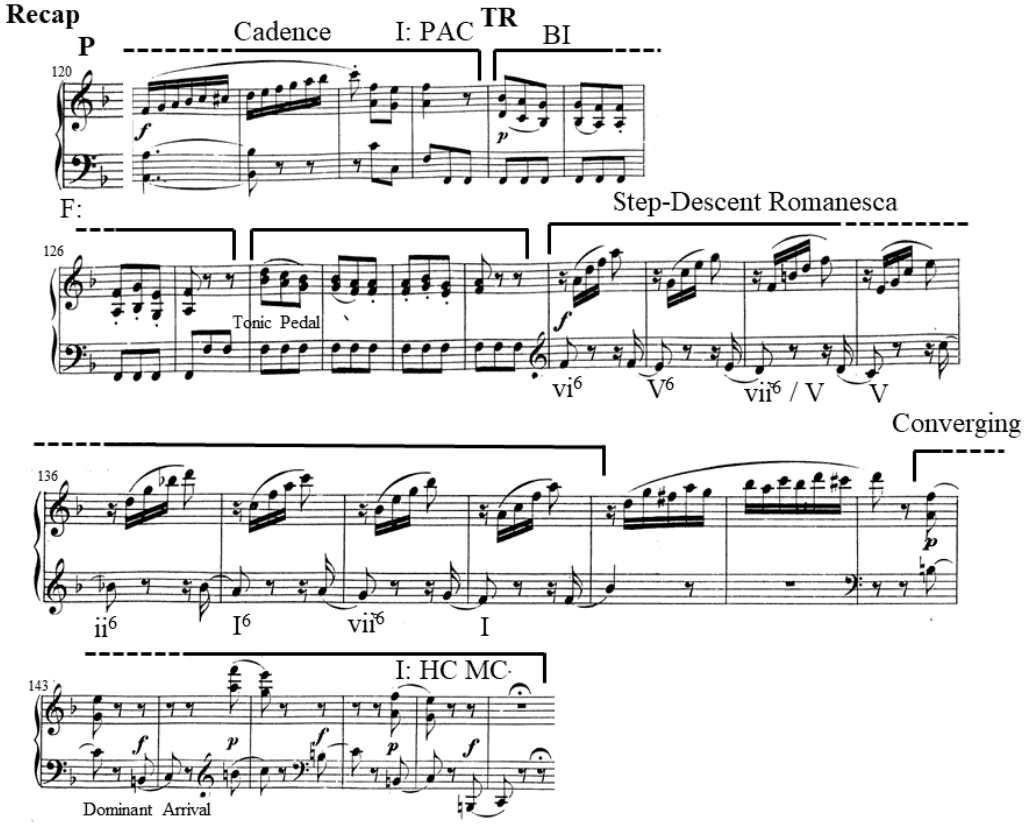


FIGURE 5.12a,b: Step-Descent Romanesca Interpretations for the Exposition (a) and Recapitulation (b) of K. 280, iii, m 120-148.

Using qualitative methods, I will demonstrate here that perceptual ambiguity or bistability in K. 280 is plausible due to 1) the overlapping features between Prinner and Step-Descent Romanesca schemata within transitional (TR) spaces, and 2) memory organization and access (i.e., LTWM expertise). I argue that the extent to which these interpretations may be *equally* available will vary with memory expertise. Similarly, the extent to which interpretations appear plausible will also vary with attention: the theoretical concept used to form the assessment or interpretation will selectively alter which memory representations are accessed. This allows for flexibility in recognition and categorization. In this way, it may be plausible to form different scale degree interpretations of a given line through focused attention. Contrastingly, it may be more challenging to form and alternate between different schemata interpretations, which requires diverted attention across multiple features simultaneously. The extent to which different interpretations may be available in either scale degrees or schemata may be unequal however, and, I argue, will likely depend on the level of expertise with the categories at hand. Therefore, for an expert, hearing a Step-Descent Romanesca in the *Presto* of K. 280 may seem entirely impossible. Such experts may, however, be able to access property simulators which are shared across schemata categories, and so be able to perceive a later modulation—but only when attending to a single line (see Figure 5.13). In this way, the overall category identity may remain a Prinner, but these experts may be able to bring online a different set of bass line simulators to remain in the previous key (i.e., perceive a Dominant Prinner).

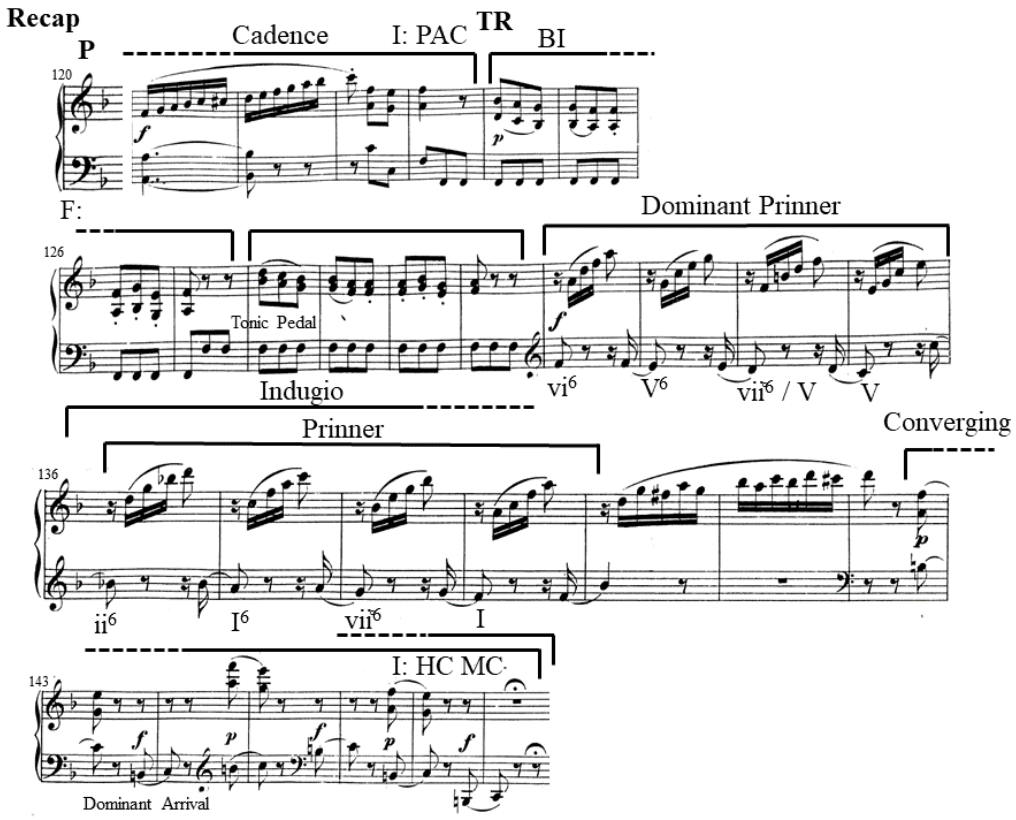


FIGURE 5.13: Monotonal Prinner Schemata Interpretation for the Recapitulatory Transition of K. 280, iii, m 120-148.

The phenomenon of interpretational flexibility—forming and assessing multiple interpretations of a musical excerpt or entire piece—is fundamental to the discipline of music theory (e.g., Cone 1977; Agawu 1994; Guck 2006). While perceptual bistability has been widely studied in the modality of vision (e.g., Brugger 1999), the investigation of the same phenomenon in the auditory modality is quite limited. Previous work has demonstrated that the effect does exist within the auditory modality, but these effects have been limited to those using artificially constructed stimuli to examine perceptual ambiguity in auditory stream segregation (so-called 'bottom-up' processes, e.g., Turgeon & Bregman 2001). Research has yet to be conducted either on bistable perception using ecologically valid stimuli, or on the ability to engage categorical knowledge in the same way that vision does (using more 'top-down' processes, e.g., to invoke

duck vs. rabbit). Previous research suggests that this type of bistability is plausible in the musical realm; interpretational flexibility has been demonstrated for tonal scale degrees in imagery for musically trained populations (Vuvan & Schmuckler 2011). The results from this study will therefore be the first to investigate perceptual bistability that engages conceptual knowledge in ecologically viable musical stimuli. In line with previous research (e.g., Hortlitz & O'Leary 1993; Firestone & Scholl 2016), the results will also seek to contribute to the understanding of bistability as arising from effects of attention (focused vs. diffuse) and memory (experience with music theory concepts).

Guiding Questions and Hypotheses

There were three primary questions that guided the development of the survey. They were:

1) Are the excerpts from K. 280, in fact, perceptually bistable?

2) If the excerpts are found to be bistable, is the formation or flexibility of interpretation dependent on which conceptual knowledge is being used to form the interpretation? In other words, do differences in auditory attending strategies and memory change the availability of interpretations or the ease of changing between two interpretations?

3) Is either forming or changing an interpretation related to the amount of expertise with relevant music theory concepts (e.g., Galant schemata)?

The study thus has one null and three primary hypotheses for each excerpt:

H0: This excerpt is not amenable to multiple interpretations, using either scale degrees or Galant schemata.

H1a: This excerpt is amenable to multiple interpretations in terms of schemata; however, one interpretation (Prinner) may be easier to hear than the other (Romanesca).

H1b: Furthermore, the availability of and ease of change between interpretations may differ between the expositional and recapitulatory versions of the excerpt (with the Romanesca more available in the recapitulation).

H2: Participants should be able to more easily form and alternate between scale degree interpretations for a single voice (soprano, bass) than for Galant schemata (which are dependent on the presence of multiple, co-occurring features).

H3: The ability to form and alternate Galant schemata interpretations may be dependent on moderate familiarity, but not a high-level expertise with Galant schemata (Expertise categories = Novice, Intermediate, Expert). Expertise should therefore be related to increased rigidity of interpretation (Ease of Change on a scale from 1 to 7, low to high), particularly for Schema interpretations, as such categories are overlearned and more likely to be automatically active during listening.

The null hypothesis is that the excerpt is not bistable, meaning that only one interpretation may be available, or that neither interpretation may be available. Hypothesis 1(a) posits that, while the excerpts may be bistable, an early modulation or Prinner interpretation may be more easily available than a later modulation or Romanesca interpretation. Hypothesis 1(b) suggests however that hypothesis 1(a) many vary based on the formal location (exposition or recapitulation). I argue that a Romanesca interpretation may be more available in the recapitulation (as it is nonmodulatory and contains a full scale from *do* to *do*) than in the exposition. This can be attributed to ‘bottom-up” factors of features in common across

categories. Essentially, an effect of representational activation of auditory imagens in the nonverbal system. As the exposition has more features in common with existing traces for modulating Prinner schemata, more of these traces are likely to be primed and active during listening. However, because this trace has some similarity to the Step-Descent Romanesca previously encountered, the traces containing these interactions may also be primed, even though there may be too few of them to give rise to activation of that schema as a whole (see Figure 5.14).[3] Therefore, one would easily be able to form the interpretation "Prinner" in the exposition, including accessing any of the relevant property or relation simulators for this category (e.g., bass line scale degrees). One might also be able to hear some similarities, albeit minimal, with an interpretation of "Step-Descent Romanesca," making the perception of Romanesca weak or unlikely, but not impossible. The availability of a Romanesca interpretation may shift for the version of the transition found in recapitulation because there are *slightly* more similarities with a Romanesca interpretation compared to the exposition. For example, the full descending major scale in the bass may help to activate more bassline traces associated with Romanaescas than the version in the exposition (see Figure 5.15).[4]

[3] For example, previously encountered exemplar traces, like those pertaining to the transition of K. 576 may be primed but not fully activated during listening. In this case, not enough simulators for the Romanesca category (bass, soprano, harmony) are active for the he category to be clearly, fully activated. Partial activation may still occur, reflecting the similarity between this instance and previously experienced instances.

[4] The full-scale interpretation is weakened by the fact that Mozart clearly creates two chunks using an octave displacement, making the reading of two consecutive Prinners quite a bit stronger. Similarly, the stronger reading of Prinner in the exposition may result in difficulties hearing any other interpretation in the recapitulation; the exposition sets the frame of reference for the second rotation.

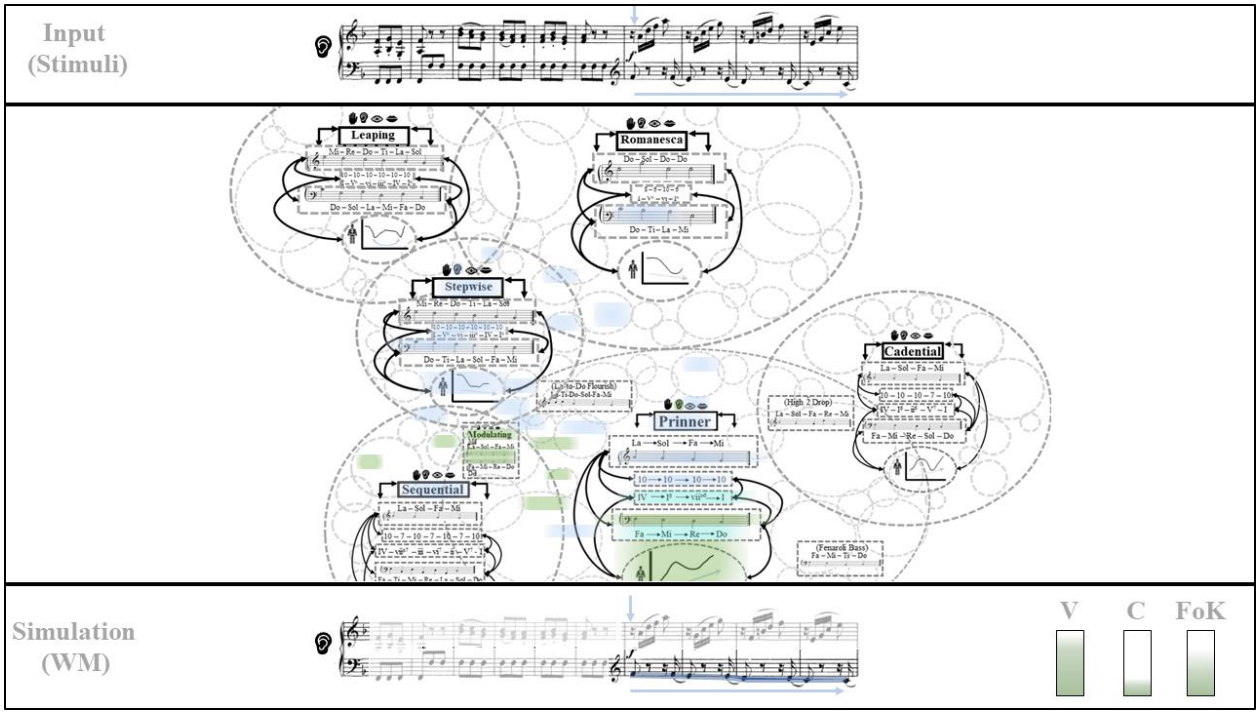


FIGURE 5.14: Direct Representational Activation of Prinner Traces and Partial Priming of Romanesca Traces in K. 280, iii, Exposition

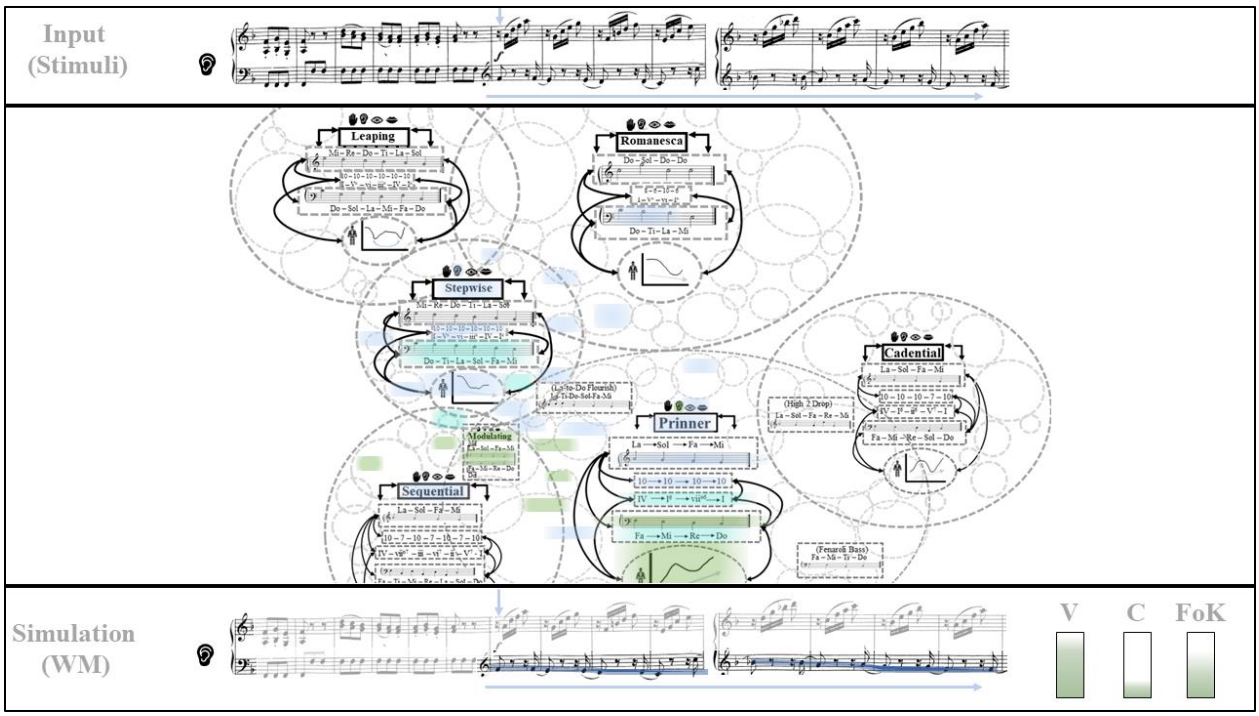
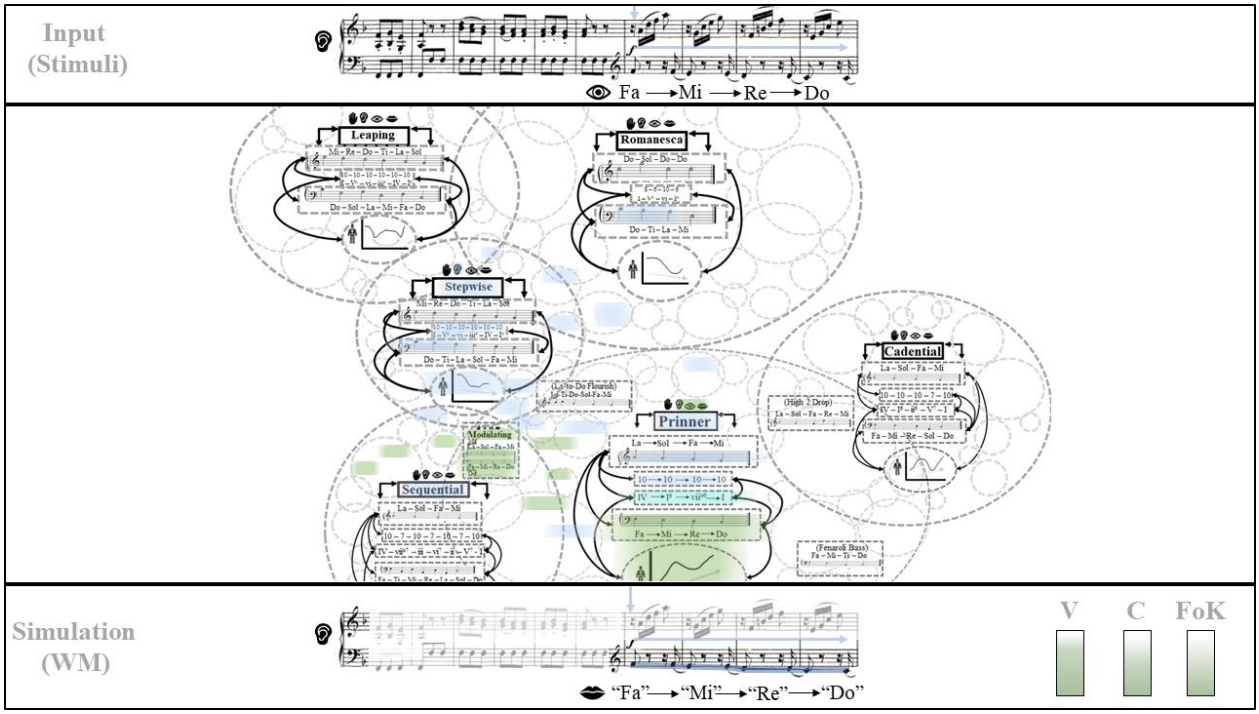


FIGURE 5.15: Direct Representation Activation of Prinner Traces and Increased Activation of Romanesca Traces in K. 280, iii, Recapitulation

The second hypothesis posits that ease of hearing and change ratings will vary with attending strategy, as a change in attention activates different memory representations. Single line attending (bass, soprano) will provide higher ease of hearing and ease of change ratings than diffuse attending (schema), as the single line attending activates a smaller pool of simulators than does diffuse attending. Essentially, the only traces that are 'essential' for a single-line simulator to be active are a smaller subset of traces that encompass previous interactions with single-line voices (e.g., sung interactions, auditory imagens, auditory-motor logogens). Other traces will certainly be active through representational, associational, and representational activation, but they do not need to be accessed and maintained in WM to make an evaluation about the goodness-of-fit for a single-line simulator (see Figure 5.16a). Hearing the line "Fa-Mi-Re-Do" is relatively simple, shown by the high vividness, control and feeling-of-knowing in WM in Figure 5.16a. Changing this interpretation to one that is "Do-Ti-La-Sol" may be difficult, but not impossible. This requires suppression of the traces that were just active and held in WM (i.e., bass line simulators). The invocation of the verbalization "Do-Ti-La-Sol" helps to activate by referential activation, auditory imagens pertaining to that simulator (see Figure 5.16b). Forming this second tonic-based interpretation may prove more difficult as fewer traces are active pertaining to this interpretation. The difficulty in forming the interpretation is represented in Figure 5.16b by the lower perceived control and feeling-of-knowing assessments in WM.

(a). Direct Representational Activation and Referential Processing of Bass Line "Fa-Mi-Re-Do"



(b). Alternation to Bass Line Interpretation "Do-Ti-La-Sol"

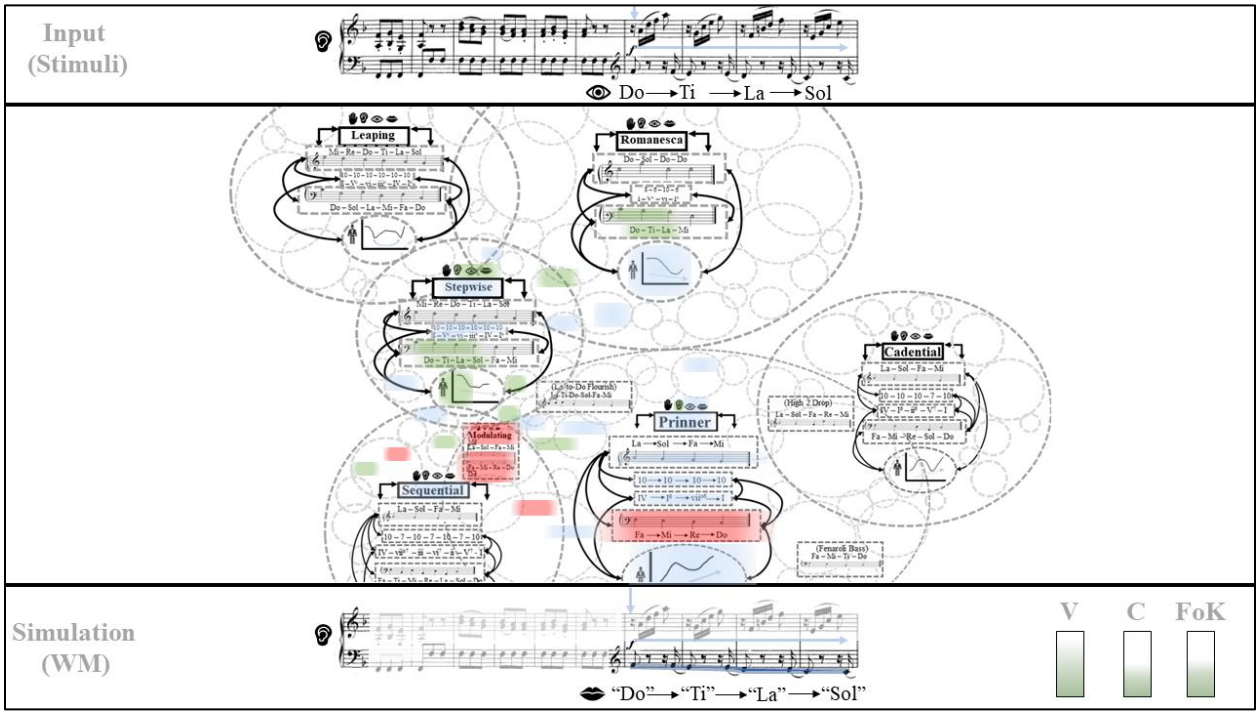
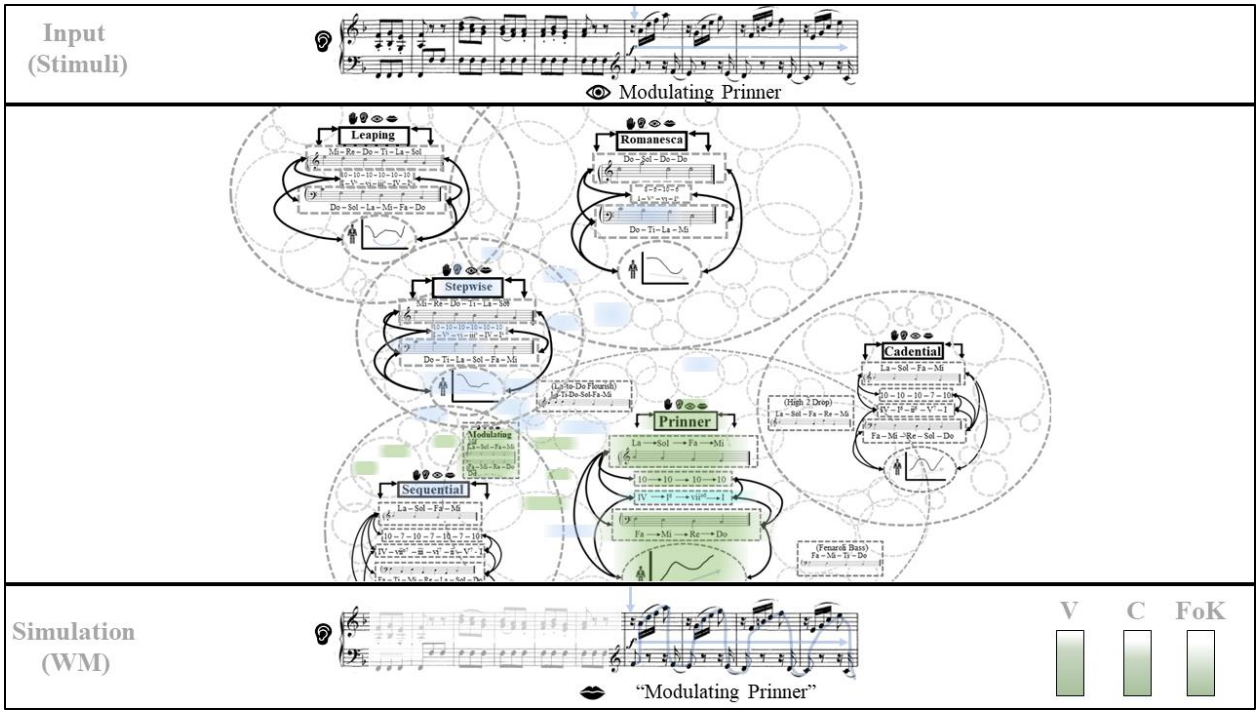


FIGURE 5.16a,b: Formation of Modulating (a) and change to Nonmodulating (b) Bass Line Interpretations in K. 280, iii, Exposition

Contrastingly, simulating and evaluating the entire category “Prinner” or “Romanesca” requires many more simulators to be active at any given time (e.g., bass line, harmony, soprano line) (see Figure 5.17a,b). While this may be relatively easy when the presented example activates through representational activation, previous traces for that category, as is the case with a Prinner interpretation (see Figure 5.17a), this is much more challenging for a Romanesca interpretation. More WM control is required for a Step-Descent Romanesca interpretation because the listener will need to suppress the many Prinner traces that automatically activate through listening. At the same time, the listener will also need to bring online (through brute force associational or referential processing) other simulators for the Romanesca (e.g., soprano voice) that may not be active through representational activation alone (see Figure 5.17b). In fact, the listener may indeed be unsuccessful at actively inhibiting Prinner simulator traces, contributing to a difficulty in hearing a Romanesca interpretation.

(a). Direct Activation of Modulating Prinner



(b). Suppression of Modulating Prinner Traces for Romanesca Interpretation

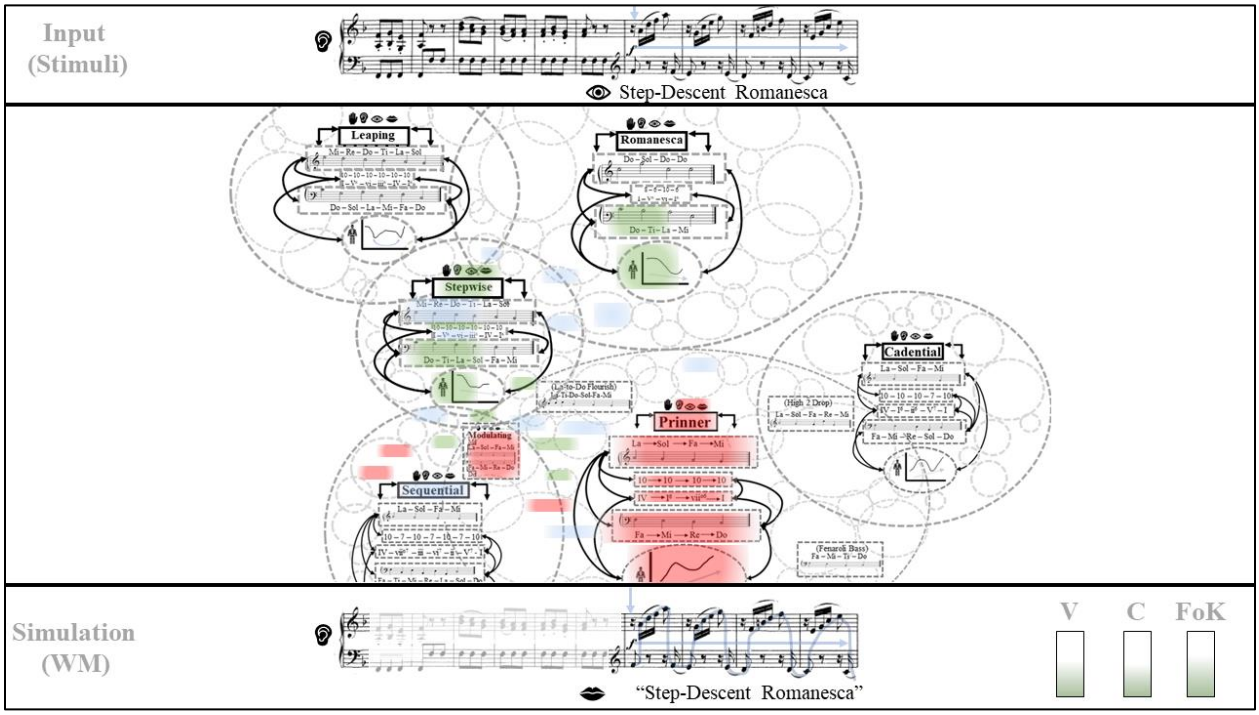


FIGURE 5.17a,b: Formation of Modulating Prinner (a) and Step-Descent Romanesca (b) Schema Interpretations in K. 280, iii, Exposition

The final hypothesis states that the extent to which different interpretations can be formed and changed will vary with expertise. Specifically, that those with intermediate level expertise will be able to form and change between two schemata interpretations far more easily than will experts. A higher level of expertise should be associated with more elaborated simulators stored in memory. As a result, higher expertise should demonstrate a rigidity of interpretation, particularly for Galant schemata, as these categories are more differentiated in memory based both on their primary features (i.e., similarity of the stimuli, or 'bottom-up' factors), and spatial-temporal location in Sonata form (i.e., so-called 'top-down' factors pertaining to the probability of activation of simulator pools over time). This is due to several factors pertaining to the structure of expert memory.

Firstly, as experts' memory is more elaborated, the *distinguishing* features of Modulating Prinners and Step-Descent Romanescas in sonata form transitions are more prominent. Because of this, representational activation will be far narrower for experts, meaning that the traces that are directly activated by listening alone will be fewer and more specific. For example, Figure 5.18 shows the formation of a schema interpretation for a hypothetical expert. Here, the traces *directly* activated by auditory input are more narrowly within the modulating Prinner sub-category, and even fewer traces within the potential feature overlap with the Romanesca are primed. Experts' memory organization therefore may simply not facilitate activation of Romanesca traces, along with suppression of Prinner traces (see Figure 5.19). This effect may also stem or be assisted by more 'top-down' features outside the stimuli. One example would be experts' experience with how schemata are employed in musical forms. Because the Modulating

Prinner tends to occur with the onset of transitional rhetoric, memory traces pertaining to the Prinner may have a higher probability of activation when an expert recognizes—overtly or not—that a transition has begun. This would not be the case with the Romanesca, which more typically occurs *after* transitional rhetoric is underway. Listeners with intermediate, as opposed to expert, levels of schema experience should show a different pattern. Since the distinguishing features ('bottom up') and distributional sensitivities ('top-down') of schemata have not yet been fully acquired and honed, listeners with some, but not a great deal, of schema expertise will likely have an easier time in both forming and alternating between the two schema interpretations.

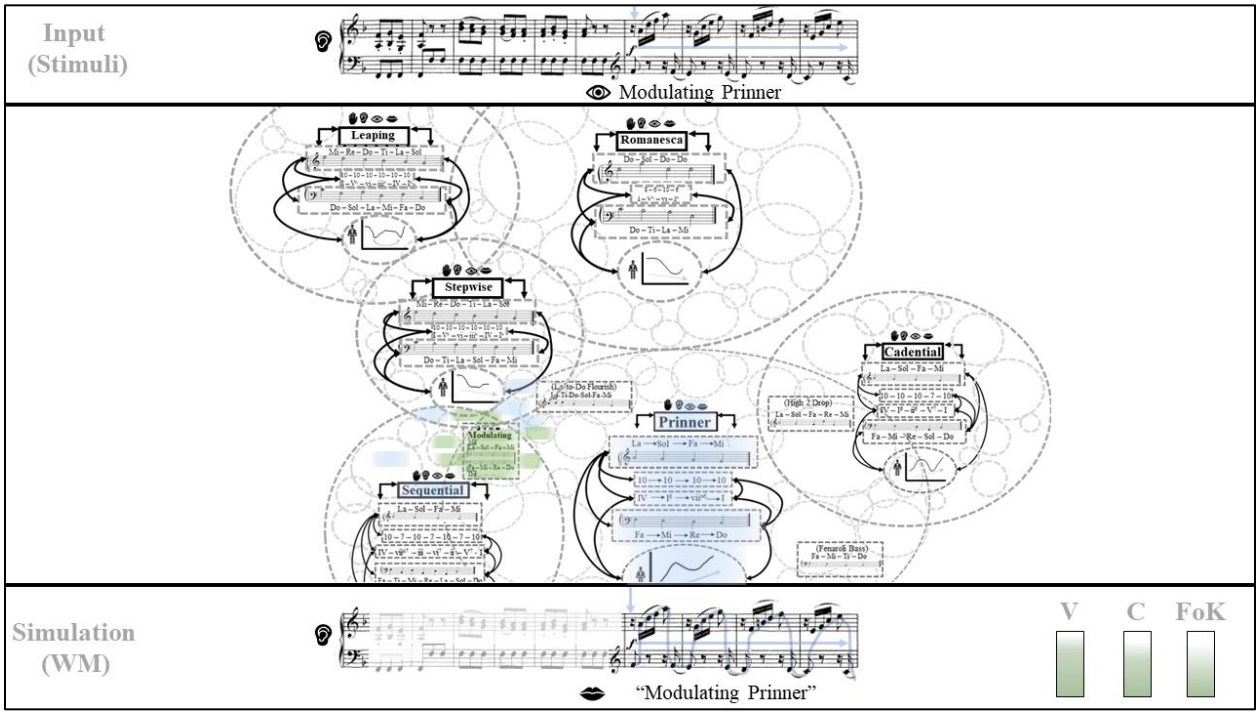


FIGURE 5.18: Direct Representational Activation of Modulating Prinner Traces of a Hypothetical Expert

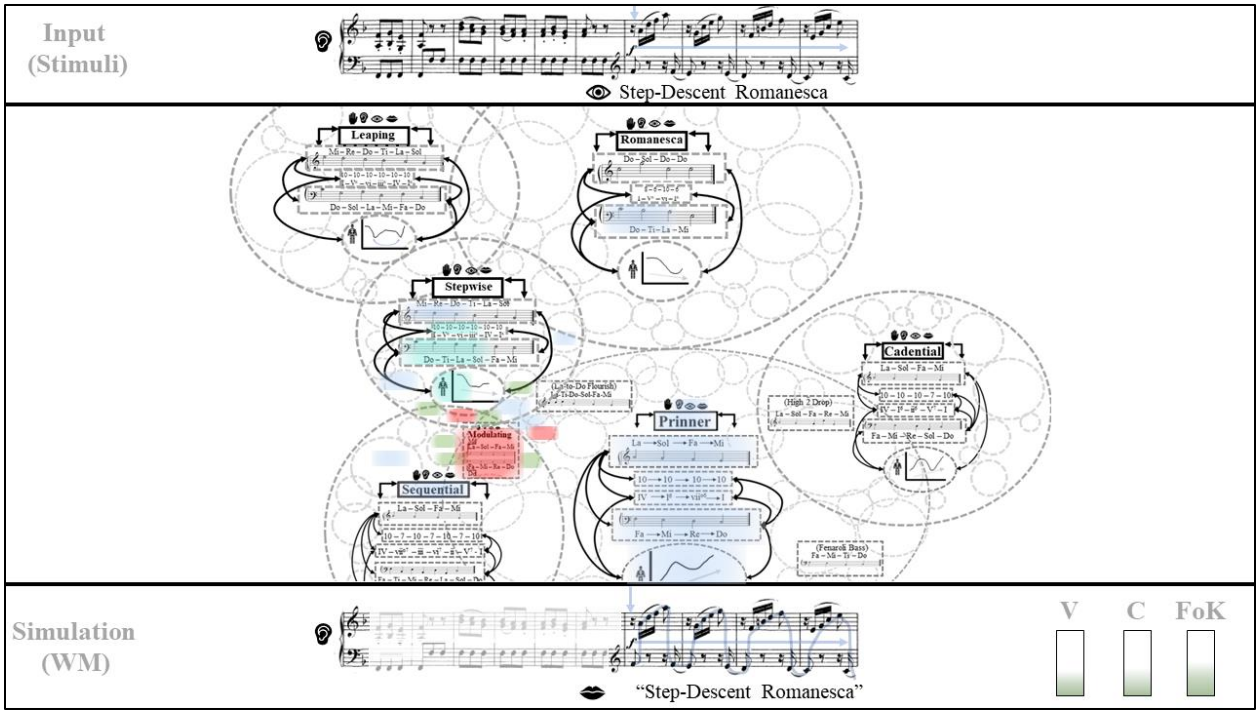


FIGURE 5.19: Suppression of Modulating Prinner Traces and Attempted Indirect Activation of Step-Descent Romanesca Traces of a Hypothetical Expert

Method

Data was collected online in Qualtrics using qualitative survey methods. This study uses methods inspired by those used in ambiguous figure perception in the domain of vision (e.g., Brugger, 1999). Participants were presented with two short musical excerpts in auditory format—the transitions from the exposition and recapitulation of Mozart's Piano Sonata no. 2, K. 280, iii—after which they were prompted to rate the ease in forming two different interpretations presented in visual format using scale degrees and/or Galant schemata labels. Following this, they were then asked to alternate between the two provided interpretations and rate the ease of change. Two types of attending strategies were investigated: 1) focused attention through scale-degree interpretations of either soprano or bass lines (i.e., a single auditory stream, see Bregman,

1990), and 2) diffuse or shifting attention, prompted through assessing an interpretation of an ordering of Galant schemata. Unlike focusing on a scale degree interpretation in a single voice, hearing a passage as a schema requires attending to multiple, co-occurring features at once. Because the latter requires more attentional resources (WM) and more active simulator traces (LTM), hearing and alternating a schema interpretation is likely more challenging. The final portion of the survey collected general information regarding the participants' music theory training (years of training), their experience with scale degrees and Galant schemata, and their familiarity with the sonata movement in the study. Taken together, the information gathered in the first part of the study addresses the first and second research questions above (i.e., whether the excerpts are bistable, and if bistability changes with conceptual prompt and auditory attending strategy), and the biographical information regarding music theory training gathered at the end of the survey addresses the final research question (i.e., whether the results observed depend on previous training and experience). The study was reviewed and approved by Northwestern's Institutional Review Board, Study IRB# 00217214.

Participants. A total of nineteen (n = 19) participated in the study. Of these, ten were professors of music theory, six were current graduate students in music theory, two were post-doctoral students in music theory, and one was a recent PhD graduate in music theory. Out of the 19 participants, a total of 3 reported having perfect pitch (all professors). Participants were recruited online through email (Music Theory and Cognition listserv in the Department of Music Studies at Northwestern, the Society for Music Theory Pedagogy Interest Group listserv, and by personal email contact) and online posts (Society for Music Theory Humanities Commons page).

Stimuli. The stimuli were two short excerpts from the expositional and recapitulatory transitions of Mozart's Piano Sonata no. 2, K. 280, iii. The audio clips were extracted from Kristian Bezuidenhout's *Mozart: Keyboard Music*, vols 8 & 9 (2016) for fortepiano using Audacity. The expositional transition included the end of the main theme through the medial caesura (see Figure 5.20) totaling 16 seconds (from 0:08-0:24 in the recording), and the recapitulatory transition contained the same adjusted material (see Figure 5.21) totaling 19 seconds (from 2:09-2:28 in the recording). Two interpretations for each excerpt were created for bass and soprano lines, as well as orderings of Galant schemata (see Figure 5.22a,b, and c).

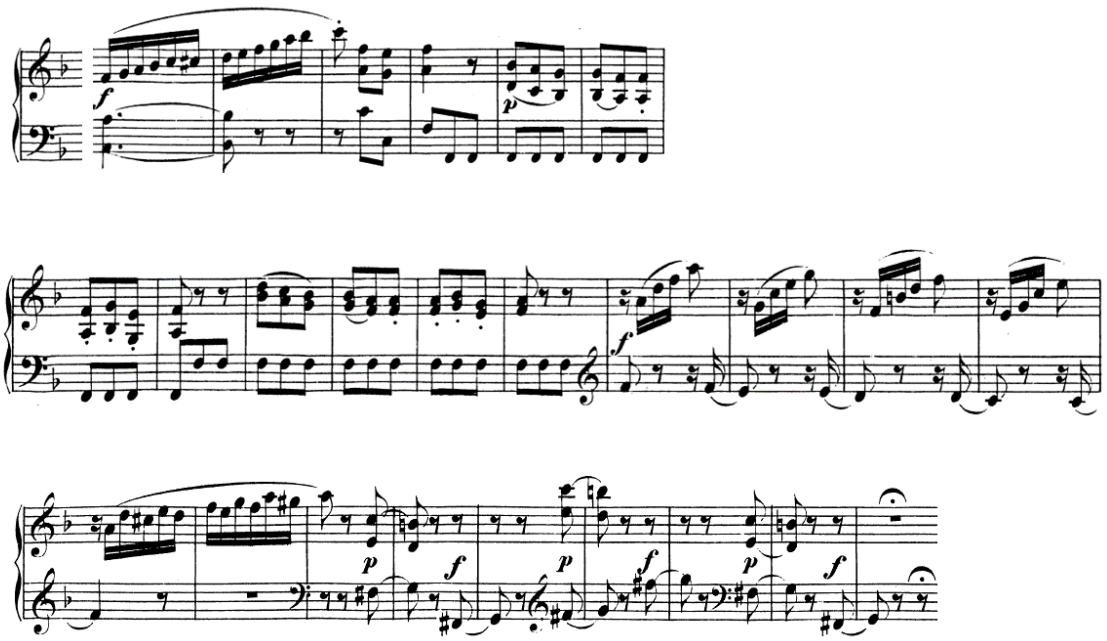


FIGURE 5.20: Expositional Transition of K. 280, iii

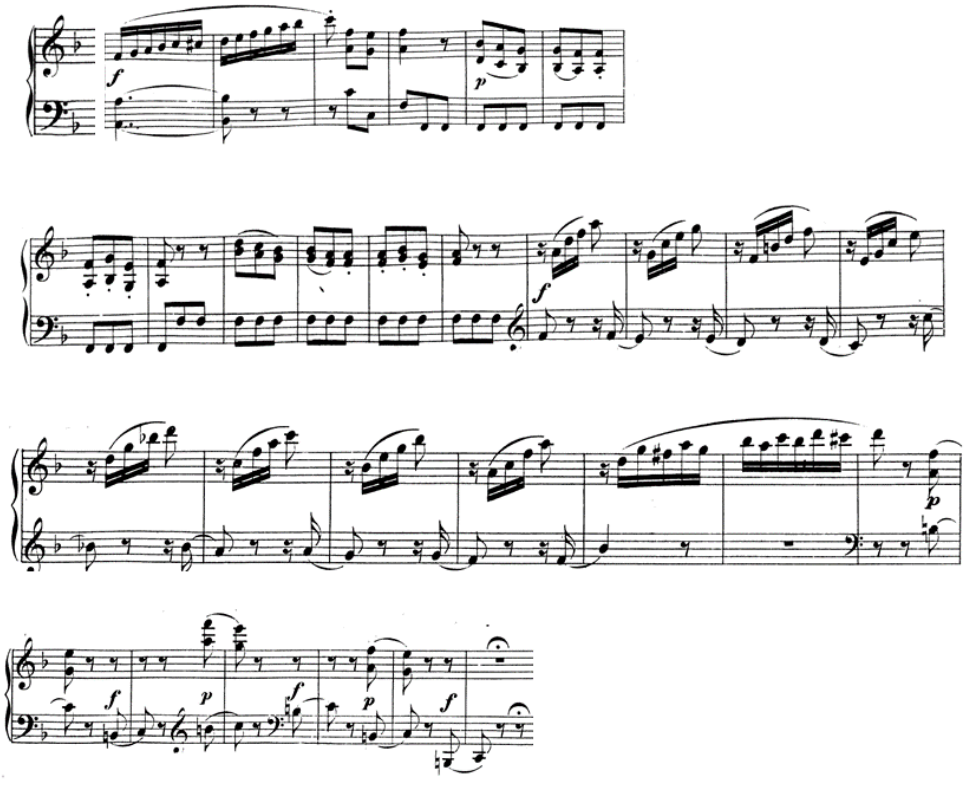
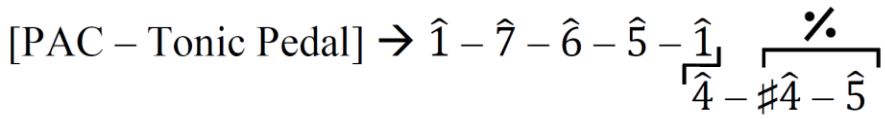
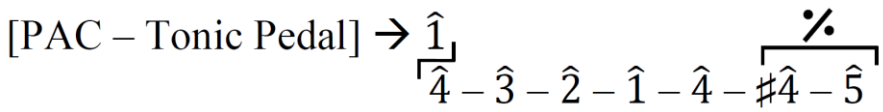


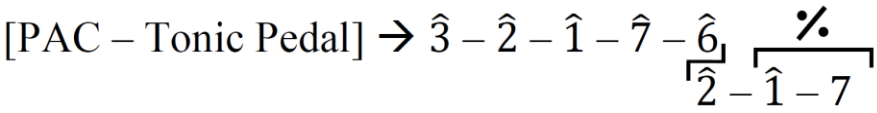
FIGURE 5.21: Recapitulatory Transition of K. 280, iii

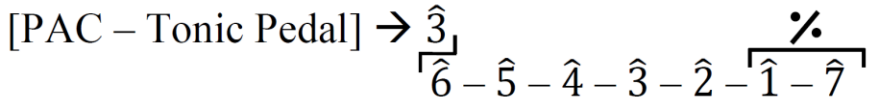
(a). Bass Line Interpretations





(b). Soprano Line Interpretations





(c). Schemata Interpretations

PAC – Tonic Pedal – Step-Descent Romanesca – Converging

PAC – Tonic Pedal – Modulating Prinner – Converging

FIGURE 5.22a,b,c: Visually Presented Interpretations for the Bass Line (a), Soprano Line (b) and Schemata (c) Interpretations of the Exposition

Procedure. The survey consisted of three parts: consent documentation and introduction, the experimental body of the survey (K. 280 ratings), and biographical information collection. At the beginning of the survey, participants read a letter of information and completed the consent documentation. They then completed an introductory training module, which introduced bistable perception through the original duck-rabbit figure drawn and presented in Jastrow (1899) (see Figure 5.23). They completed the task using this picture. Firstly, they were asked to form the interpretation RABBIT, and rate the ease of forming this interpretation on a Likert scale from 1 (cannot see at all) to 7 (can clearly see it). Following this, they were presented with a second interpretation, DUCK, and completed the same procedure. Participants were then prompted to view the image continuously and rate the ease of changing the interpretation from RABBIT to DUCK on a Likert scale from 1 (cannot alternate interpretations) to 7 (can easily alternate interpretations). Lastly, participants replicated the task in a practice session with a musical stimulus—the modulating transition from the exposition of Mozart's Piano Sonata no 7, K. 309, i. They were asked to rate the ease of hearing two bass line interpretations (see Figure 5.24) on a Likert scale from 1 (cannot hear at all) to 7 (can easily hear), as well as the ease of changing these two interpretations (1-7).

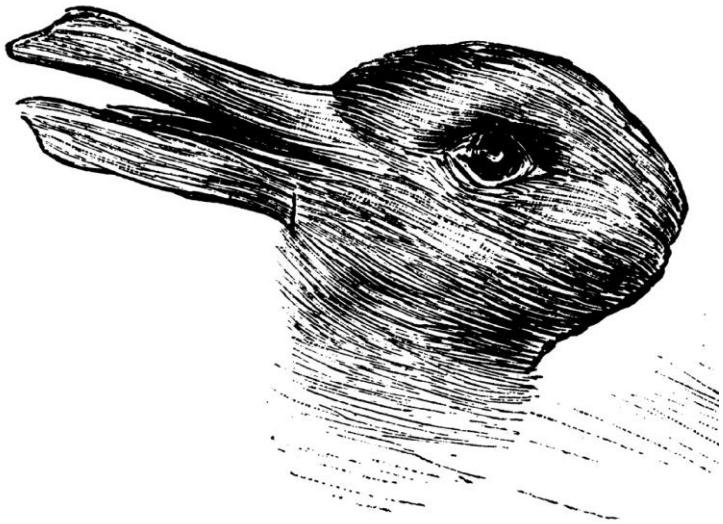


FIGURE 5.23: Duck-Rabbit Ambiguous Figure from Jastrow (1899)

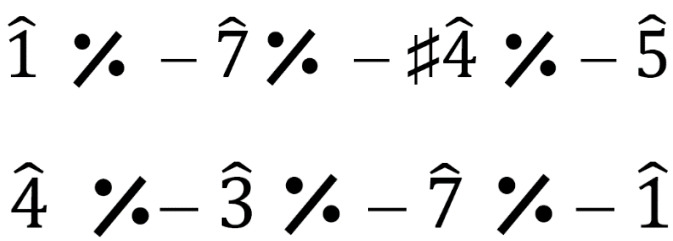


FIGURE 5.24: Bass Line Interpretations for K. 309 used in the Survey Introduction

The experimental body of the survey gathered ease of hearing and ease of change ratings for the two excerpts from the Presto of K. 280. The survey was organized into two main blocks (exposition, recapitulation), which contained three blocks for each attentional condition (bass line, soprano line, schemata). Each attentional condition (bass, soprano, schemata) contained three trials each, one rating for each interpretation, and one rating for the ease of change between the two presented interpretations. This resulted in a total of 18 trials. The presentation order of the excerpts was randomized such that the exposition or recapitulation blocks could be presented in either order, the attentional condition blocks (bass, soprano, schema) could be presented in any order, and within each condition, the order of presentation of each interpretation trials (e.g., Prinner, Romanesca) was randomized. Before starting the first set of ratings in either the exposition or recapitulation, participants were prompted to listen to the excerpt without any interpretation provided. Once within a given block (e.g., Recapitulation) and condition (e.g., Schema), participants completed both ease of hearing and ease of change ratings before moving onto the next condition (e.g., bass line). Once all three conditions were completed (bass, soprano, schema), they moved onto the next main block (e.g., exposition). Participants were provided with an audio clip of the excerpt for each rating and could listen as many times as they liked. For the ease of change ratings, participants were encouraged to listen to the excerpt multiple times and to change interpretations on each successive listening.

The final part of the survey gathered various biographical information that was used to create expertise designations (see *Data cleaning and pre-processing* below). Participants were prompted to provide their education level, number of years of music theory and aural skills experience (both training and teaching), expertise in scale degree hearing and with Galant schemata, their familiarity with the excerpt (hearing and analysis), if they had perfect pitch, and what types of strategies they used for forming and changing interpretations in the task they had just completed.

Data cleaning and pre-processing. The raw data exported from Qualtrics was cleaned in python using pandas and numpy libraries. The data was aggregated into long format, which resulted in a total of 228 observations for ease of hearing ratings (DV1) and 114 observations for ease of change ratings (DV2). Data pre-processing included the labeling of block and trial orders (e.g., Sonata Order 1: exposition, recapitulation. Sonata Order 2: recapitulation, exposition), as well as the creation of schemata expertise designations from biographical data provided on schema familiarity and schema hearing. Schema familiarity involved a categorical response: 'Completely unfamiliar', 'Somewhat familiar', 'Familiar', 'Very familiar', and 'I am an expert (analysis, playing, composition, etc.).' Schemata hearing was measured on a Likert scale from 1 (Strongly disagree to 7 (Strongly agree) where participants rated the statement "I can hear Galant Schemata while listening." Expertise groups for schemata therefore involved an aggregate of both responses. For the categorical response, those responding as experts were classified as experts, those familiar and very familiar were classified as intermediate, and those somewhat familiar and completely unfamiliar classified as novices. The ordinal schemata hearing response was used to verify these classifications. To remain in the intermediate level, participants required a schemata hearing value between 3-6, otherwise they were placed in the novice group. Those

with the categorical response ‘somewhat familiar’ whose schema hearing level was 3 or higher were placed in the intermediate group. This resulted in three participants with ‘somewhat familiar’ responses to be placed in the intermediate group ($M_{Schema\_Hearing}$ = 4.53, sd = 0.45) and two participants with ‘somewhat familiar’ responses to be placed in the novice group ($M_{Schema\_Hearing}$ = 1.50, sd = 0.51). One participant was manually placed in the expert group whose categorical response was ‘very familiar’ and schema hearing value was 6.7 because this participant was specifically contacted by the author for their expertise in Galant schemata.

A summary of the schemata expertise group data is shown below (see Figure 5.25). Each group did not appear to differ greatly on their self-reported ability to hear scale degrees, indicated by the scale degree hearing ratings in the figure below. Average schema hearing levels scaled by group, with experts having the highest ratings (n = 5, M_{Schema} = 6.38, sd = 1.10, M_{Scale_Degree} = 6.64, sd = 0.46), novices having the lowest (n = 4, M_{Schema} = 1.25, sd = 0.43, M_{Scale_Degree} = 5.50 sd = 1.51), and intermediates in between (n = 9, M_{Schema} = 4.85, sd = 1.10, M_{Scale_Degree} = 6.01, sd = 0.65).

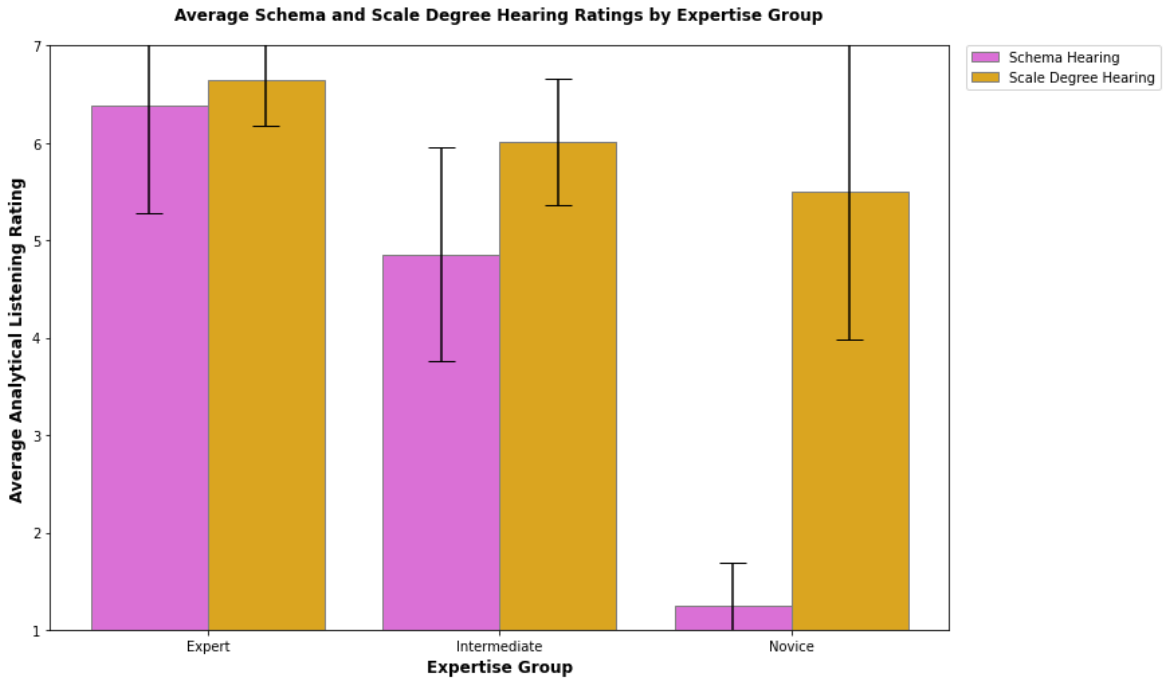


FIGURE 5.25: Average Schema Hearing and Scale Degree Hearing Ratings for Each Expertise Group

Results

The survey design resulted in a four-way mixed ANOVA design, with 2*3*2 factorial design for the within-subject factors, and three levels of the between-subject factor. The within-subject factors included two levels of sonata section (exposition, recapitulation), three levels of attended feature (bass, soprano, schema) and two levels of modulation type (Prinner, Romanesca). The between-subject factor included the three levels of expertise group (expert, intermediate, novice). Upon exploratory analysis of the results, one participant in the expert group was dropped due to internal inconsistencies in their responses.

Ease of hearing ratings (DV1). A four-way mixed ANOVA was performed in R using the rstatix package to evaluate the effects of sonata section, attended feature, modulation type and expertise on the ease of hearing ratings. The dataset included sphericity violations for the within-subject factors as assessed by Shapiro-Wilk's test of normality for all conditions except for ratings in the exposition in the soprano condition for the Romanesca modulation type ($p = 0.109$). As a result, p-values using Greenhouse-Geisser corrections (where applicable) were used. There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances.

The analysis revealed main effects of expertise group, $F(2, 15) = 6.81$, $p = 0.008$, and attended feature, $F(1.94, 29.10) = 11.37$, $p = 0.0002$, $\varepsilon = 0.970$.Post-hoc Tukey HSD tests revealed that both Expert ($M = 5.13$, $sd = 2.12$) and Novice ($M = 3.56$, $sd = 2.17$) groups differed significantly, $p < 0.001$, 95% C.I. = [-2.44, -0.696], and that Intermediate ($M = 5.40$, $sd = 1.66$) and Novice ($M = 3.56$, $sd = 2.17$) groups differed significantly, $p < 0.001$, 95% C.I. = [-2.63, -1.05] (see Figure 5.26). For attended feature, a post-hoc Tucky HSD test showed that the main effect of feature comes from a statistically significant difference between the Bass ($M = 5.45$, $sd = 1.65$) and Soprano ($M = 4.35$, $sd = 2.12$) features, $p = 0.003$, 95% C.I. = [-1.89, -0.313] (see Figure 5.27). The main effect for sonata section was nonsignificant, $F(1, 15) = 0.009$, $p = 0.926$, indicating that ratings for the exposition ($M = 4.93$, $sd = 2.04$) were equal to the recapitulation ($M = 4.90$, $sd = 2.05$). The main effect for modulation type was also nonsignificant, $F(1, 15) = 4.41$, $p = 0.053$, indicating that ratings for the Prinner ($M = 5.21$, $sd = 1.99$) did not differ significantly from those for the Romanesca ($M = 4.62$, $sd = 2.06$).

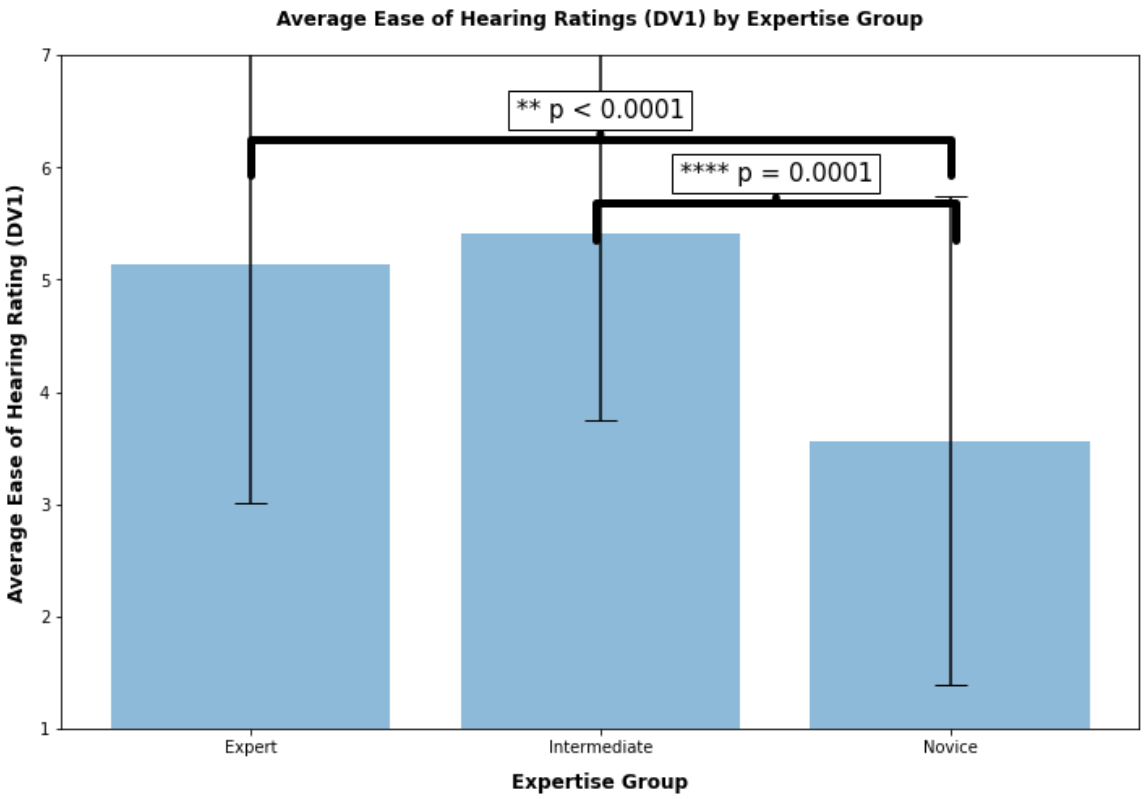


FIGURE 5.26: Average Ease of Hearing Ratings (DV1) by Expertise Group

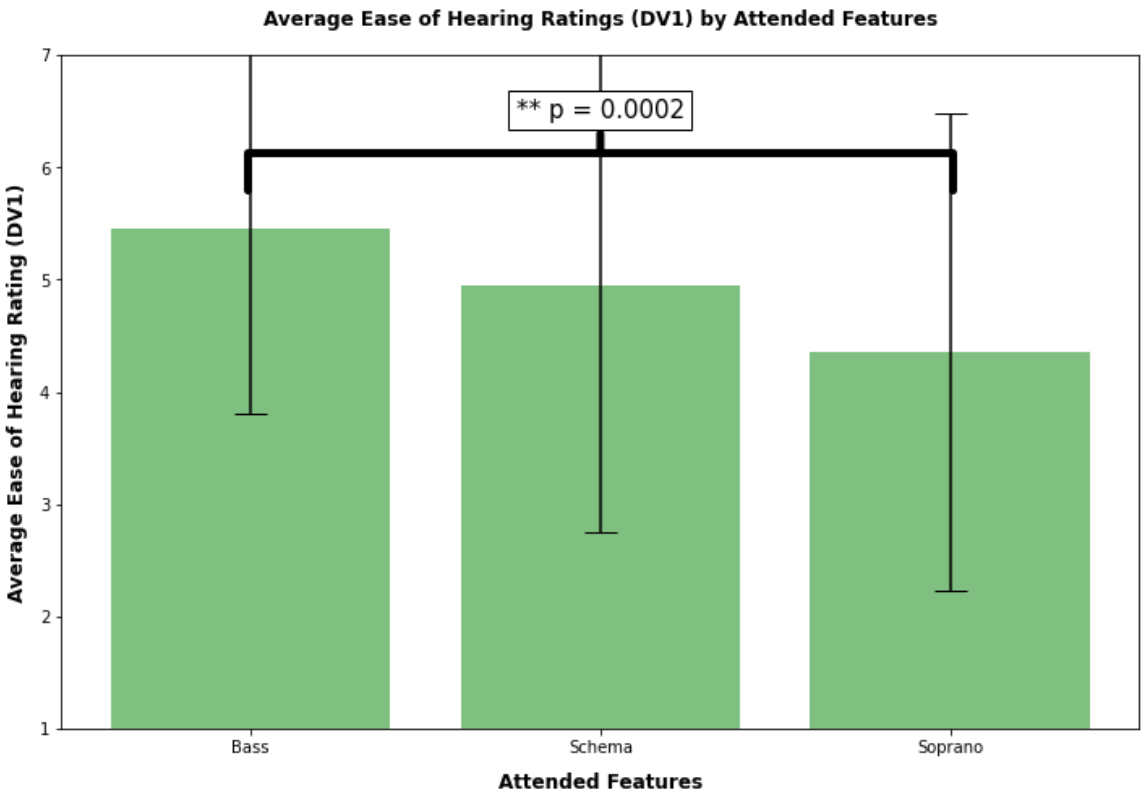


FIGURE 5.27: Average Ease of Hearing Ratings (DV1) by Attended Feature

The four-way mixed ANOVA analysis also showed four significant two-way interactions, but no significant three- or four-way interactions. There were two significant interactions that involved within-subject factors only, and two that involved interactions between the within- and between-subject factors (i.e., effects of expertise). For the within-subject factor interactions, the analysis revealed a significant interaction of sonata section by attended feature, $F(1.64, 24.61) = 6.34$, $p = 0.009$, which was corrected for sphericity violations ($\varepsilon = 0.820$). A post-hoc Tukey HSD test revealed that the Bass ($M = 5.60$, $sd = 1.47$) and Soprano ($M = 4.12$, $sd = 2.17$) differed significantly only for the recapitulation, $p = 0.005$, 95% C.I. = [-2.58, -0.364] (see Figure 5.28). There was also a significant interaction between attended feature and modulation type $F(1.95, 29.10) = 4.75$, $p = 0.017$, which was corrected for sphericity violation

using the Greenhouse-Geisser correction (ε = 0.970). A post-hoc Tukey HSD test revealed that only for the Schema attending condition, the Prinner (M = 5.73, sd = 1.86) was rated higher than the Romanesca (M = 4.16, sd = 2.23), p = 0.001, 95% C.I. = [-2.53, -0.602] (see Figure 5.29). The two-way interaction between sonata section and modulation type was nonsignificant, F(1, 15) = 3.45, p = 0.083. Both Prinner and Romanesca were equally available in the exposition ($M_{Prinner}$ = 5.33, sd = 1.93 | $M_{Romanesca}$ = 4.54, sd = 2.09) and recapitulation ($M_{Prinner}$ = 5.08, sd = 2.06 | $M_{Romanesca}$ = 4.71, sd = 2.04).

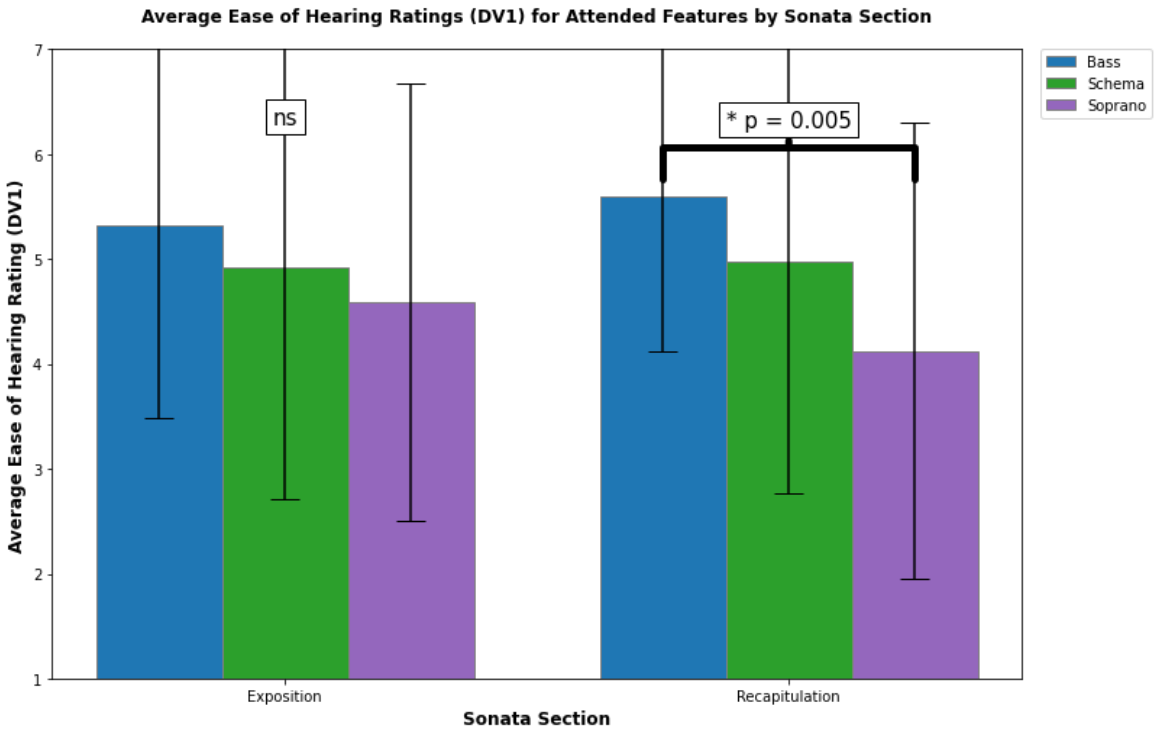


FIGURE 5.28: Average Ease of Hearing Ratings (DV1) for Attended Features by Sonata Section

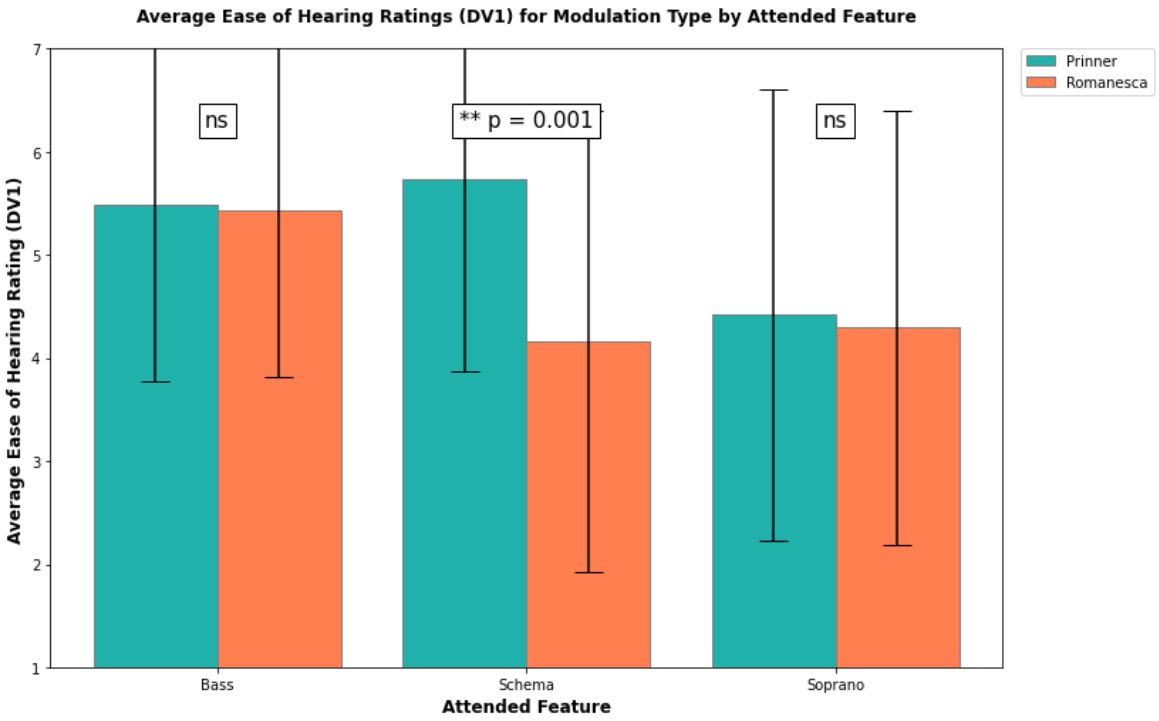


FIGURE 5.29: Average Ease of Hearing Ratings (DV1) for Modulation Type by Attended Feature

Lastly, the analysis revealed two significant interactions involving the between-subject factor, expertise group. Firstly, there was a significant two-way interaction between expertise group and attended feature, $F(3.88, 29.10) = 3.12$, $p = 0.031$, which was corrected for sphericity violations ($\varepsilon = 0.970$). Post-hoc Tucky HSD test showed that for the Novice group only, the Bass ($M = 4.83$, $sd = 2.07$) and Soprano ($M = 2.51$, $sd = 1.62$) conditions differed significantly from one another, $p = 0.005$, 95% C.I. = [-4.03, -0.619] (see Figure 5.30). Secondly, there was a significant two-way interaction of expertise group and modulation type, $F(2, 15) = 4.25$, $p = 0.034$. A post hoc Tucky HSD revealed that that for the Expert group only, the Prinner ($M = 6.24$, $sd = 1.49$) was rated higher than the Romanesca ($M = 4.02$, $sd = 2.08$), $p < 0.0001$, 95% C.I. = [-3.15, -1.28] (see Figure 5.31). The two-way interaction between expertise group and

sonata section was nonsignificant, F(2, 15) = 0.803, p = 0.467, showing that ratings for the exposition and recapitulation did not differ between groups: Expert (M_{expo} = 5.30, sd = 2.08 | M_{recap} = 4.96, sd = 2.17), Intermediate (M_{expo} = 5.42, sd = 1.61 | M_{recap} = 5.39, sd = 1.72), Novice (M_{expo} = 3.40, sd = 2.19 | M_{recap} = 3.72, sd = 2.19). All three- and four-way interactions were nonsignificant.

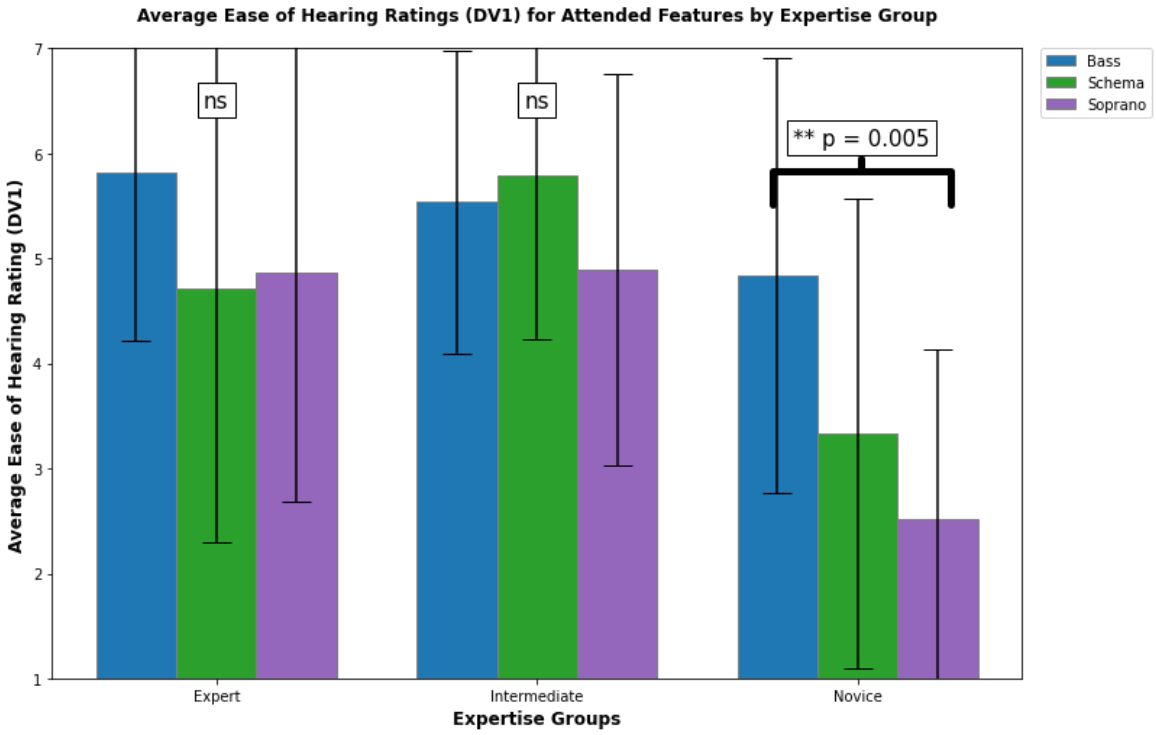


FIGURE 5.30: Average Ease of Hearing Ratings (DV1) for Attended Features by Expertise Group

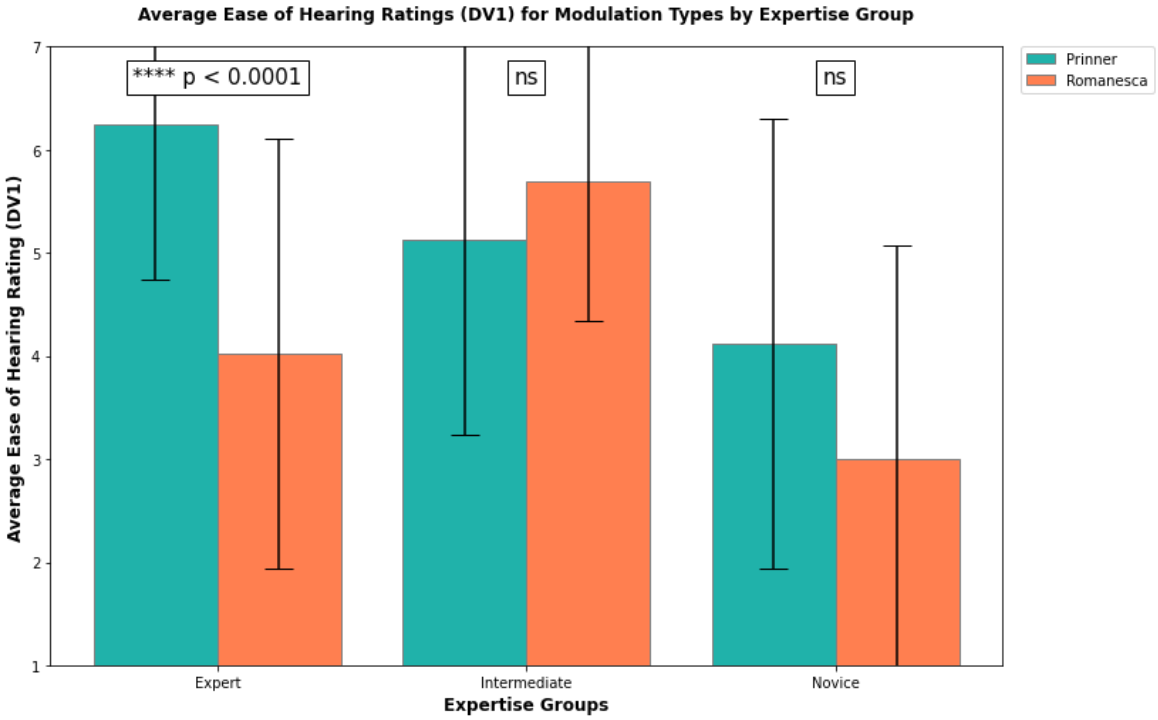


FIGURE 5.31: Average Ease of Hearing Ratings (DV1) for Modulation Type by Expertise Group

Ease of change of interpretation ratings (DV2). A second three-way mixed ANOVA was performed to examine the effects of expertise group (expert, intermediate, novice), sonata section (exposition, recapitulation), and attended feature (bass, soprano, schema) on ease of change ratings (DV2). The dataset included sphericity violations for the within-subject factors as assessed by Shapiro-Wilk's test of normality for all conditions except for bass and schema attending conditions in recapitulation ($P > 0.05$). As a result, p-values using Greenhouse-Geisser corrections were used for the analysis. There was homogeneity of variances ($p > 0.05$) as assessed by Levene's test of homogeneity of variances. The analysis revealed significant main effects of expertise, $F(2, 15) = 4.54$, $p = 0.029$, and feature, $F(1.51, 22.64) = 4.81$, $p = 0.026$. Post hoc Tukey HSD revealed both the expert and intermediate groups differed from the novice group: Expert ($M = 3.57$, $sd = 2.06$) and Novice ($M = 1.88$, $sd = 1.50$), $p = 0.00273$, 95% C.I. = [-2.89, -0.511], and Intermediate ($M = 4.49$, $sd = 1.82$) and Novice ($M = 1.88$, $sd = 1.50$), $p < 0.0001$, 95% C.I. = [-3.68, -1.55] (see Figure 5.32). For attended feature, a similar pattern to DV1 was observed where the Bass ($M = 4.11$, $sd = 1.99$) was rated higher than the Soprano ($M = 3.09$, $sd = 2.06$), but a post hoc Tuckey HSD revealed that this was nonsignificant ($p = 0.113$). The main effect of sonata section was nonsignificant, $F(1, 15) = 0.262$, $p = 0.616$, ($M_{Exposition} = 3.79$, $sd = 2.20$ | $M_{Recapitulation} = 3.52$, $sd = 1.97$). All two-way interactions were nonsignificant.

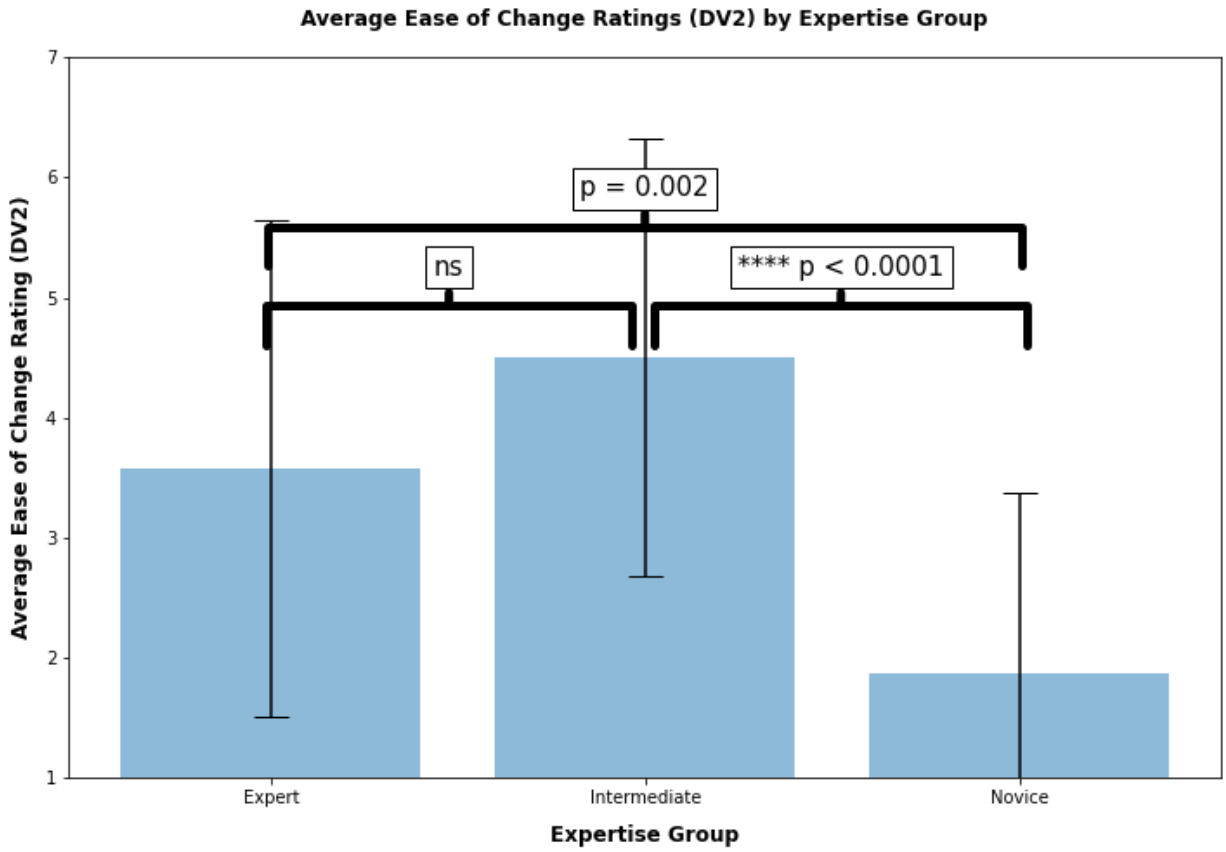


FIGURE 5.32: Average Ease of Change Ratings (DV2) by Expertise Group

Excerpt familiarity and expertise. The relationship between excerpt familiarity and expertise group was examined in order to ascertain if the observed effects of expertise group in the analyses above could be attributed to familiarity with the excerpt. It is plausible that higher familiarity with the piece could lead to rigidity in interpretation formation and ease of change. Two questions in the survey were designed to gather different types of familiarity ratings. First, one question regarding familiarity with the excerpt where participants selected one of three response types: “This was the first time I have heard this”, “I have heard this before”, “I am very familiar with this piece”. A second question gathered information on whether or not participants had analyzed the piece before. For this, participants selected one of three response types pertaining to whether they had analyzed the piece: “No, not at all”, “No, not at all”, or “Yes, extensively”. **Table 5.1** shows the response breakdown by expertise group for both excerpt familiarity and excerpt analysis. For the expert group, all participants responded that they had heard the piece before, three indicated that they had not analyzed the piece and two indicated that they had somewhat analyzed the piece. For the intermediate group, four had never heard the piece before, four had heard it before, and one participant indicated that they were very familiar with the piece. For excerpt analysis, seven in the intermediate group indicated that they had never analyzed the piece, while the remaining two indicated that they had somewhat analyzed it. In the novice group, three participants indicated that they had never heard the piece before, and one indicated that they had heard it before. All participants in the novice group indicated that they had never analyzed the piece. No participants indicated that they had extensively analyzed the piece. From this data, familiarity categories were created (see **Table 5.2).** These categories were created using the following criteria:

- Low Familiarity: Indicated “This was the first time I have heard this” for excerpt familiarity and “No, not at all” for excerpt analysis.
- Mid Familiarity: Indicated “I have heard this before” for excerpt familiarity and “No, not at all” for excerpt analysis.
- High Familiarity: Indicated “I have heard this before” for excerpt familiarity and “Yes, somewhat” for excerpt analysis OR “I am very familiar with this piece” and for excerpt familiarity and “Yes, somewhat” for excerpt analysis.

TABLE 5.1:

Question Type	Responses	Expert	Intermediate	Novice
Excerpt Familiarity	“This was the first time I have heard this”	0	4	3
	“I have heard this before”	5	4	1
	“I am very familiar with this piece”	0	1	0
Excerpt Analysis	“No, not at all”	3	7	4
	“Yes, somewhat”	2	2	0
	“Yes, extensively”	0	0	0

TABLE 5.2:

Excerpt Familiarity Categories	Expert	Intermediate	Novice
Low	0	4	3
Medium	3	3	1
High	2	2	0

An ordinal logistic regression was performed in order to see if familiarity category (low, medium, high) was able to significantly predict expertise group (novice, intermediate, expert). The predictive (AIC=455.23) model was not a better fit than the null model (AIC = 451.82, X^2 = 0.59, p = 0.7423), Pseudo R^2 = 0.001(McFadden). This indicates that excerpt familiarity is not predictive of expertise group. Therefore, the observed results of expertise group above can be attributed to generalized schemata expertise, and not to familiarity with the excerpt itself.

Discussion

Here, I will contextualize the findings in light of the original hypotheses and theoretical framework developed in this dissertation. Firstly, we can safely reject the null hypothesis:

H0: This excerpt is not amenable to multiple interpretations, for either scale degree interpretations or Galant schemata.

Both excerpts were amenable to multiple interpretations for both scale degree lines and Galant schemata. Regarding the first hypothesis:

H1a: This excerpt is amenable to multiple interpretations in terms of schemata; however, one interpretation (Prinner) may be easier to hear than the other (Romanesca).

H1b: Furthermore, the availability of and ease of change between interpretations may differ between the expositional and recapitulatory versions of the excerpt (with the Romanesca more available in the recapitulation).

Both hypotheses 1a and 1b were rejected as there was no main effect of modulation type, nor an interaction between modulation type and sonata section. This demonstrates that the excerpt demonstrated relatively equal perceptual bistability between interpretations (Prinner, Romanesca), and that this did not vary between the exposition and recapitulation.

H2: Participants should be able to more easily form and alternate between scale degree interpretations for a single voice (soprano, bass) than for Galant schemata (which are dependent on the presence of multiple, co-occurring features).

The second hypothesis was partially confirmed. The significant main effect of attended feature showed that overall, bass lines and Galant schemata were equally perceptible, but that the soprano line was much more difficult to hear. The significant interaction of sonata section and attended feature showed that this effect was largely due to differences between the bass and soprano line ratings (DV1) in the recapitulation. I interpret this effect as arising from difficulty in forming interpretations in the exposition in general. However, without a main effect of sonata section, it is difficult to confirm this. The significant interaction of attended feature and modulation type also confirms the second hypothesis: early and late modulation interpretations are equally available when attending to the bass voice; however, when switching to a multi-feature attending strategy needed to assess Galant schemata interpretations, Prinner and Romanesca become much less equally available.

H3: The ability to form and alternate Galant schemata interpretations may be dependent on moderate familiarity, but not a high-level expertise with Galant schemata (Expertise categories = Novice, Intermediate, Expert). Expertise should therefore be related to increased rigidity of interpretation (Ease of Change on a scale from 1 to 7, low to high), particularly for Schema interpretations, as such categories are overlearned and more likely to be automatically active during listening.

The final hypothesis regarding expertise was confirmed in multiple ways. The significant main effects of expertise for both DV1 and DV2 showed that the Intermediate group had the highest ease of hearing and ease of change ratings, confirming that they were more easily able to hear both interpretations, and were much more able to alternate between interpretations. The low ease of change (DV2) ratings in both novice and expert groups are particularly interesting because these lower ratings can be attributed to different potential causes. For the novices, the

low ease of change (DV2) ratings can be attributed to a *lack* of category representations for Galant schemata. Contrastingly, the expert groups' lower ease of change ratings can be attributed narrow activation of highly elaborated simulators in memory, resulting in a rigidity of interpretation. For the intermediate group, there are enough Galant schemata representations to selection from, but the sparser (less elaborated) nature of these traces means that alternating simulator bases is much easier.

The interaction of expertise and attended feature for DV1 indicates that for the novice group only, scale degree hearing for the soprano line (compared to the bass line) was very low, whereas for both intermediate and expert groups, this difference was nonsignificant. This result is particularly interesting given that the novice group reported relatively high levels of scale degree hearing ability (see Figure 5.25 above). I interpret this result in light of schemata expertise: as both intermediate and expert groups had higher familiarity with Galant schemata, I argue that they were likely able to automatically 'fill in' the misaligned soprano line information in imagery (LTM to WM) as soprano line simulators were more likely to automatically come online during listening (see Figure 5.33). When the tight association between bass and soprano line simulators have not been acquired, as is the case in the novice group, they need to rely solely on the perceptual input—or representational activation alone—of existing scale degree lines, which results in a lower availability of interpretation in this voice.

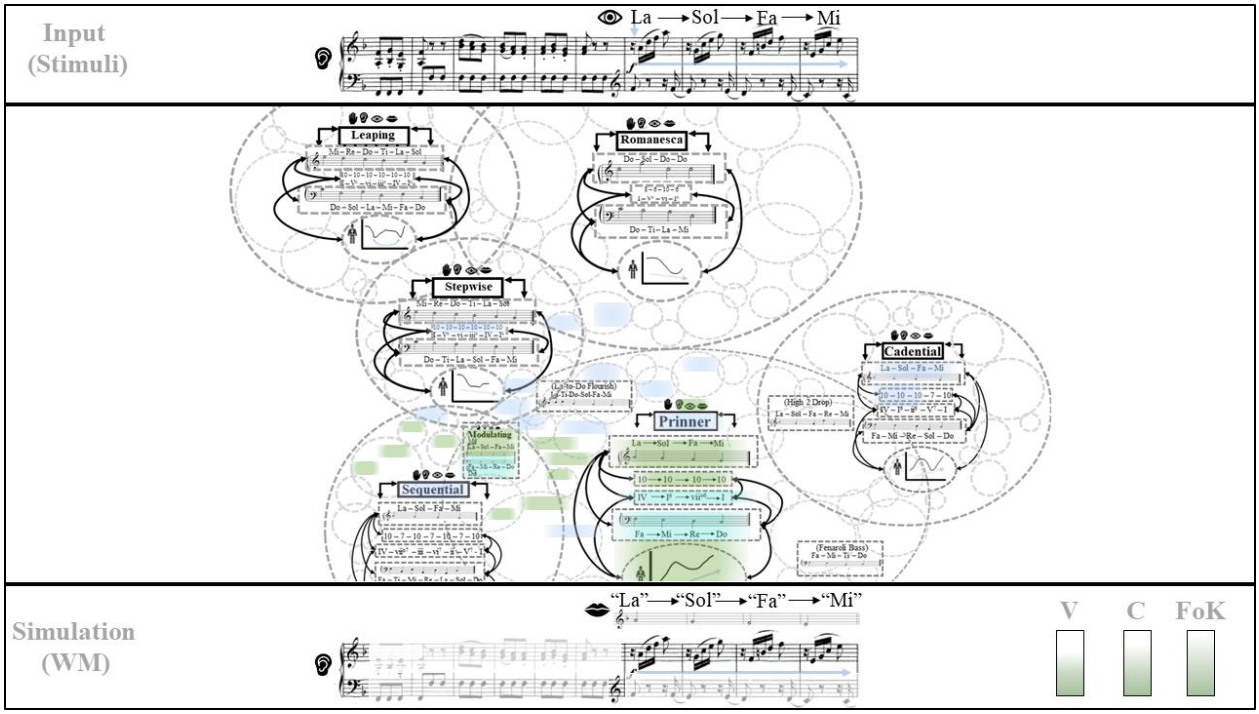


FIGURE 5.33:

Lastly, the interaction of expertise group and modulation type for DV1 shows that for the expert group only, the Romanesca hearing is more difficult compared to a Prinner hearing. Contrastingly, and as predicted, Prinner and Romanesca categories are much more equally available for the intermediate group. As there is no significant three-way interaction between expertise group, feature, and modulation type, this indicates that the availability of both Prinner and Romanesca interpretations for the Expert group did not differ between attending types (see Figure 5.34). Therefore, even when only attending to the bass line, the experts' Galant schema knowledge (i.e., automatic activation of simulator pools through associational and referential processing) exerts an influence on the simulators that are available from LTM to WM. In this way, hearing a later modulation (or no modulation at all, i.e., Romanesca, or dominant Prinner) is far less compelling than hearing an earlier modulation (i.e., modulating Prinner), regardless of whether the experts use a focused (i.e., bass line) or diffuse (i.e., schema) attending strategy.

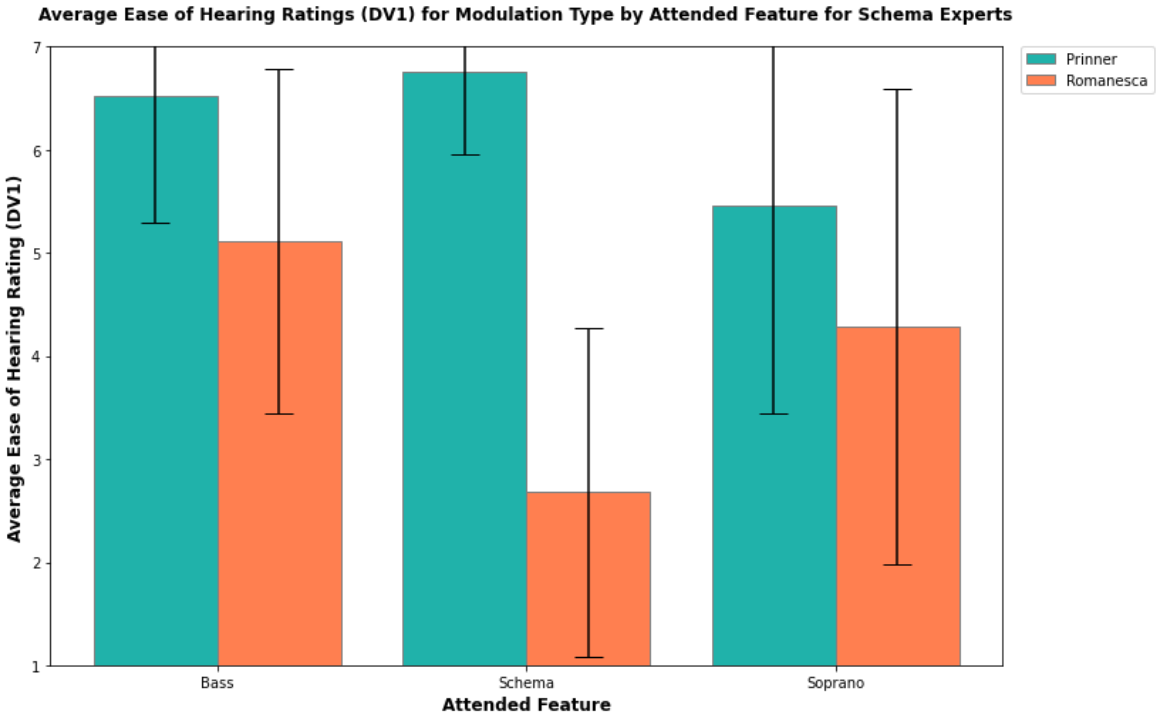


FIGURE 5.34: Average Ease of Hearing Ratings (DV1) for Modulation Type by Attended Feature for Schemata Experts

As predicted, it appears that this is due to both bottom-up and top-down factors. One of the experts' responses[5] for how they formed their interpretations suggests exactly this:

> In addition, on a more reflective level, of course there are cues for the activation of the Modulating Prinner schema in a spot like this (e.g., surface activity that's speedy, harmonic modulation, connecting to Converging Cadence pattern at the end)—the top-down cue of "this is 3-over-1 becoming immediately 6-over-4 in the new key" is probably the strongest factor of all, which doesn't really leave much

[5] This participant was in the Mid-level excerpt familiarity group.

> room for ambiguity. In terms of the formal grammar and even the surface affordances, there's simply no way that this is a (home-key) Prinner riposte.'

Therefore, it very much appears as if the ‘surface cues’ (bottom-up) facilitated direct, representational activation of traces pertaining to the modulating Prinner. Even when presented in a contextually limited context, such cues were strong enough to elicit explicit recognition of the formal location of the excerpt through associational and referential processing. Along with the limited amount of distributional (order-over-time) information available in the short excerpt (top-down), this severely limited the categorization options for the schema here, resulting clearly in a reduction of ease of hearing ratings for Romanesca or late-modulation Prinner, and lower ease of change ratings overall for expert listeners.

Lastly, an examination of the interpretation formation strategies revealed important roles for imagery and subvocalization involving both musical and verbal content. **Table 5.3** provides a breakdown of the strategies reported by participants for each expertise group (expert, intermediate, novice), and for all participants together (sum). Most categories were equally represented across groups. The most common strategies were imagining or subvocalizing scale degrees or solfege during listening (n = 19 responses). Some participants reported that they sang or spoke scale degrees or solfege out loud during listening (n = 6 responses) or sung or verbalized out loud without scale degrees or solfege while listening (n = 4 responses). Some participants reported speaking or singing between listenings (n = 4), and very few reported playing an instrument (n = 2). These responses suggest that imagery and verbalization play vital roles in simulation during interpretation formation.

TABLE 5.3: Interpretation Formation Types by Expertise Group

Interpretation Formation Type	Expert	Intermediate	Novice	Sum
I have no idea, I could just hear it!	1	1	2	4
I imagined scale degrees or solfege as I listened	3	6	4	13
I imagined or sung/spoke the interpretation between listenings	1	1	2	4
I sang or spoke (without scale degrees or solfege)	1	1	2	4
I subvocalized (silent rehearsal, spoken or sung) scale degrees or solfege while listening	2	4	2	8
I spoke or sang scale degrees or solfege while listening	1	4	1	6
'I played an instrument	2	1	0	3
Other (please specify)	1	0	0	1

Summary and Conclusions

In this chapter, I have presented a case for music theory LTWM as control over interpretation formation and modification in perception. Previous research into ambiguous figure perception reveals important contributions of both top-down and bottom-up factors. The analytical case of Mozart's piano sonatas demonstrates a potential for perceptual ambiguity between Modulating Prinners and Step-Descent Romanescas in Sonata Form transitions due to overlapping features. The qualitative survey functions to support this claim by demonstrating that the transition of K. 280, iii, is indeed perceptually bistable. However, this study also reveals effects of top-down factors on the ease of forming and changing an interpretation: the effects of expertise in particular demonstrates that the extent to which interpretations are available and can be modified depends partially on the level of experience with Galant Schemata categories.