University of Cincinnati

Date: 6/27/2024

I, Clair Hoang Khuong Nguyen, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Theory.

It is entitled:

Beyond Mickey Mousing: Audiovisual Rhythm in Animated Film Musicals and Multimedia

Student's name: **Clair Hoang Khuong Nguyen**

This work and its defense approved by:

Committee member: Samuel Ng, Ph.D.

Committee chair: Cristina Losada, Ph.D.

Committee member: Christopher Segall, Ph.D.

48210

Beyond Mickey Mousing:
Audiovisual Rhythm in Animated Film Musicals and Multimedia

A dissertation submitted to the
Graduate College
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

In the Division of Composition, Musicology, and Music Theory
of the College-Conservatory of Music

June 27, 2024

by
Clair Hoang Khuong Nguyen

B.M., University of South Florida, 2014
M.M., University of South Florida, 2016
M.A., Hunter College, City University of New York, 2018

Committee Chair: Dr. C. Catherine Losada, PhD
Committee Member: Dr. Christopher Segall, PhD
Committee Member: Dr. Sam Ng, PhD

Abstract

The filmic experience of animated multimedia can be deeply understood if sync points are identified for their functions. My methodology combines Sergei Eisenstein's film editing concepts with music theories of rhythm and meter to contribute a *sync point analysis* framework for studying rhythmic audio (music) and visual (image) interactions in multimedia such as American and Japanese animated musicals as well as mobile rhythm games. In this framework, music shows *what* image is synchronized to, while image shows *how* music is synchronized. I argue that applying the study of film rhythm to music-theoretical frameworks of phrase structure, formal function, and rhythm and meter helps understand the creation and consumption of films. While music theory and film theory each have specialized terminology and limitations, I connect shared concepts between fields to develop an encompassing interdisciplinary framework.

This dissertation focuses on identifying sync points in animated multimedia. At the same time, I seek to push even the most on-the-nose synchronicity, commonly known as mickey-mousing, beyond its status as a mere plot accessory or visual gimmick. My use of "synchronicity" refers to precise moments when image and sound closely and rhythmically mirror each other in fluctuating periodicities. I place audio and image on equal grounds to explore how music-theoretical and visual film-editing techniques define the sync point functions. I argue that functional sync points feature visually marked moments that support musical functions, whether formal, rhythmic, metric, or cadential. That is how sync points pinpoint *audiovisual* relationships that are *reciprocal*, equal, and functional when audio and image are perceived together.

In the initial chapters, I develop a framework for *sync point* analysis that (1) differentiates synchronous from nonsynchronous audio-image relations and (2) assigns labels to describe sync

point functions. This system's originality lies in delineating specific musical and filmic features for analysis. I devise a plugin formula n[x]-[y] as the label, whereby *n* stands for rate of rhythmic occurrence, container [x] stands for music accent type, and container [y] stands for visual accent type. Case studies drawing from Disney Renaissance films, Disney's *Fantasia 2000*, Japanese *anime* series, and more demonstrate the utility of *sync point* analysis across a diverse repertory of animated film musicals and music videos. Once the n[x]-[y] opens pathways to analysis, larger-scale observations can be made of form and cadential action.

The sync types become building blocks for *sync trains*, or a visual map of sync points and connective in-between actions. Sync trains highlight emergent rhythms formed across wide stretches of music and give insight into formal and narrative interpretations. Then, case studies of "Close to Gray" from the *Hatsune Miku: Colorful Stage!* series expands the framework to include video games and exemplify how sync point analysis can also help understand felt synchronicities in a haptic (touch-based) dimension. Such multimedia utility of the framework demonstrates the robust analytical capabilities of relating musical harmony to rhythmic film editing as well as the benefit for contributing interdisciplinary systems of music analysis to a body of existing visual analysis methodologies.

Acknowledgements

When I first began my PhD studies in 2018, I was simply navigating from various points A and B. Yet, since then, I've come to learn that all the meaningful events between points A and B are what truly make the journey. In the process of completing this dissertation, I realized that my philosophical underpinnings inform my analytical framework, vice versa, both of which are shaped by my dynamic experiences in and out of school. These experiences are only possible because of the very special people I've met along the way.

First, I would like to thank my dissertation committee for all your guidance, support, and advice throughout the years. Thank you for also being exceptional graduate assistant supervisors. Through your direction, my growing confidence in teaching with clarity and communication found its way into my writing habits. Very special thanks to Dr. Catherine Losada for your guidance as my dissertation advisor, graduate teaching supervisor, and professor. I could not have made it this far as a person, student, scholar, or teacher without your invaluable feedback. Thank you for believing in my dissertation topic when it formed in transformational theory class and for your unending insights in improving my frameworks for new frontiers in academia.

Thank you to my committee first chair, Dr. Chris Segall. I did not expect my review of *Hollywood Harmony* in readings class would underlie my shift from Russian Music studies to film music analysis, and I greatly appreciate the chance to deep-dive into Prokofiev and Eisenstein's audiovisual collaborations in our Russian music theory independent study. Thank you to my committee second chair, Dr. Sam Ng, for opening my eyes to the music theories of rhythm and meter that are so integral to my dissertation. Your openness to watch my roughly synchronized video of "Porky's Railroad" and Shostakovich's third movement of the Piano Concerto No. 2, Op. 102 laid the foundation for my sync trains and related terminology.

Many thanks to Dr. Gary Weissman, Dr. Stephen Meyer, and Dr. Daniel Batchelder for laying the foundations for my Film and Media Studies certificate through independent studies tailored for this dissertation. I would also like to thank all my previous teachers who directly opened my pathways to music theory: Dr. Ciro Scotto and Dr. Anne-Marie Scotto, Dr. Jill Brasky, Dr. Ya-Hui Cheng, Dr. Philip Ewell, and Professor Poundie Burstein.

Thank you to all my friends, colleagues, and family—all my aunts and uncles—for your friendship and support throughout the years. I feel like every one of you is graduating with me as I take a step forward. Thank you to my CCM friends and family (you know who you are!) for a wonderful experience at the University of Cincinnati and for showing me the importance of having a strong cohort. Thank you to my Genshin at Ten family (you know who you are!) for strengthening my understanding of friendship. Even though we (ironically) don't usually meet at our planned 10:00 pm, we were always there for each other at times of great changes and growth.

Thank you to my awesome and inspiring sister, Catherine, who is always there to chat and talk some sense into me, and for all the last-minute tech-related and music notation saves. I treasure all the times we spent and adventure together. I am proud to have a dissertation that touches on the film and video game music industries that you are a part of.

Finally, I extend my deepest thanks to my parents, Hal and Mai-Huong Nguyen, who always showed me the light at the end of my many tunnels—and patiently waited for me there. This dissertation would not be possible without your unyielding love, determination, and sacrifices ever since Catherine and I were fledgling homeschoolers. Thank you for always traveling with me to conferences and listening to my papers and presentations. Most of all, thank you for believing in me and my path in pursuing a PhD in Music Theory, even when I did not believe in myself and even when it felt like an eternity!

To my late grandparents who left too early but will forever live in my heart.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1: Background

The filmic experience of animated multimedia can be deeply understood if sync points are identified for their functions. Therefore, this dissertation responds to the current state of sound film, where technological advances can produce (1) precise audiovisual correlations and (2) highly mediatized audiences that consume audiovisual products. This dissertation treats *audiovisual products* as any media form that features synchronized sound and image. From a creation-oriented *poietic* standpoint, an editor's precise audiovisual correlations and production choices can manufacture a sense of fidelity and hyperrealism communicated through immediate rhythmic interactions of images and sound. From a reception-oriented *esthesic* angle, audiences who are entrained to the vast body of film media are at least subconsciously expectant of certain recurring film rhythm patterns that arise in specific genres, including but not limited to action films, animation, and filmed musicals.[1]

Sound film was preceded by silent film. To provide music for silent film, an ensemble accompanied image with preexisting music, including classical music.[2] Rick Altman traces

[1] The terms *poietic*, creation-oriented analysis, and *esthesic*, reception-oriented analysis, are drawn from Jean-Jacques Nattiez's writings on music semiotics. His writings focus on analyzing for meaning and structure, including the English-horn solo from Wagner's *Tristan and Isolde*, act III. For more, see the most recent English translation in Jean-Jacques Nattiez, *Musical Analyses and Musical Exegesis: The Shepherd's Melody in Richard Wagner's Tristan and Isolde*, trans. Joan Campbell Huguet (New York: University of Rochester Press, 2021), doi:10.1017/9781800103634.

[2] In Daniel Goldmark's discussion of piano accompaniment for silent animation, he mentions the manual *Motion Picture Moods for Pianists and Organists,* indicating a tradition of using familiar melodies for accompanying silent cinema. These manuals often sorted pre-existing music into categories based on mood and affect. For more, see Daniel Goldmark, *Tunes for 'Toons: Music and the Hollywood Cartoon* (Los Angeles: University of California Press, 2005), 14.

sound film's origins to the early 1900s, beginning with silent photography.[3] Sound film's potential found its way on vaudeville productions, recording opera performances, newsreels, and cartoon reels, eventually adapting developing technology's "sound-on-film recording process" around 1926.[4] Altman and Buhler attribute the *Jazz Singer* (1927) as an early example of sound film, and what is considered the first talkie utilized Lee De Forest's sound-on-film system to achieve unprecedented audiovisual synchronization in both speech and singing scenes.[5] This technology made possible sound film musicals, like those scored by Max Steiner, and synchronized cartoons such as Disney's 1928 *Steamboat Willie*, known for "mickey-mousing."[6] The cartoons provided a new avenue of sonic and slapstick expressions. According to Susan Smith, animation allows for exaggerated performance and is "capable of stretching, compressing, and changing shape in response to the rhythms and patterns of music to a degree that is impossible in the live-action musical."[7] It is this very expressiveness of animation that this dissertation seeks to explore through a rhythm-oriented methodology.

The Soviet cinema tradition oversaw some of the earliest theories on sound film and audiovisual synchronicity. Soviet director Sergei Eisenstein's transitional writings from silent to sound film provide grounds for connecting disparate topics of film montage, animation, music, and audiovisual synchronicity.[8] In 1928, he, with fellow directors Pudovkin and Alexandrov,

---

[3] Rick Altman, "Introduction: Sound/History," in *Sound Theory, Sound Practice*, ed. Rick Altman, 113-125 (New York: Routledge, 1992).
[4] Ibid.
[5] Ibid., 120–121. See also James Buhler, *Theories of the Soundtrack*, 1 (New York: Oxford University Press, 2019).
[6] Roy Prendergast, *Film Music: A Neglected Art* (New York: W.W. Norton, 1992).
[7] Susan Smith, "The Animated Film Musical," in *The Oxford Handbook of the American Musical*, eds. Raymond Knapp, Mitchell Morris, and Stacy Wolf (New York: Oxford University Press, 2011): 167-178.
[8] Eisenstein himself has written on his fascination of Disney animations; a sentiment shared by Prokofiev. Fiona Ford and Lea Jacobs have related Eisenstein's audiovisual concepts to cartoons in their work. For more, see:

Fiona Ford, "The Film Music of Edmund Meisel (1894–1930)," PhD dissertation, (University of Nottingham, 2011), http://eprints.nottingham.ac.uk/id/eprint/12271.

Lea Jacobs, *Film Rhythm after Sound: Technology, Music, and Performance* (Oakland: University of California Press, 2015). Prior to her book, Jacobs also published an analytical article on describing music and image relations and Eisenstein's filmmaking methodologies for the film *Ivan the Terrible*. See also, Lea Jacobs, "A Lesson

penned the 1928 "Statement on Sound" critically discussing audiovisual synchronicity.[9] Of the three, Eisenstein developed the oft-cited audiovisual score marking synchronous points between scenes from *Alexander Nevsky* (1938) and Prokofiev's original score.

Presently, there are only a handful of "image-sound" graphing techniques inspired by Eisenstein's early *audiovisual score*.[10] Eisenstein had a musical way of theorizing his montage techniques, with words such as "rhythmic," "metric," and "tonal," all terms found in this dissertation. The term *audiovisual* often circulates in studies of modern media as well as the works of Eisenstein. For this dissertation, I adapt a similar understanding as Birger Langkjær, who broadly generalized "audiovisual" studies as involving "*vertical synchronous interaction* and *horizontal narrative implications*."[11] Kia Afra (2015) and Katya Ermolaeva (2020) are among scholars who go in-depth on Eisenstein's *audiovisual* theories. Afra's comprehensive commentary on Eisenstein's vertical montage theories bring forth three important factors: that (1) Eisenstein conceptualizes synchronous audiovisual film in a "temporal dimension," (2) that his theories moved from an early focus on dialectic conflict to that of synthesis, synchronicity, and synesthesia, and (3) that his theoretical shift was complicated by his application of older dialectic terms to a more organicist outlook.[12] In this vein, my dissertation's application of Eisenstein's terms rekindles the very issue of applying dialectic terms to the more synthetic framework of

---

with Eisenstein: Rhythm and Pacing in Ivan the Terrible, Part I," in *Music and the Moving Image* 5, no. 1 (2012): 24-46, doi:10.5406/musimoviimag.5.1.0024.

[9] Statement reprinted in Sergei Eisenstein, *Film Form: Essays in Film Theory*, trans. Jay Leyda (New York: Harcourt, Inc., 1977), 257–60.

[10] Rick Altman, "Visual Representation of Film Sound as an Analytical Tool," in *The Oxford Handbook of Film Music Studies*, ed. David Neumeyer (Oxford: Oxford University Press, 2013).

[11] Birger Langkjær, "Audiovisual Styling and the Film Experience: Prospects for Textual Analysis and Experimental Approaches to Understand the Perception of Sound and Music in Movies," in *Music and the Moving Image* 8, 2 (2015): 35–47, doi:10.5406/musimoviimag.8.2.0035.

[12] Kia Afra, "'Vertical Montage' and Synaesthesia: Movement, Inner Synchronicity, and Music–Image Correlation in Alexander Nevsky (1938)," in *Music, Sound and the Moving Image* 9, 1 (2015): 33-61. doi: 10.3828/msmi.2015.2.

sync point analysis. However, I will demonstrate through case studies and construction of larger graphs that the terms, once mapped with music-theoretical principles, can help create parameters for *projecting* recurring audiovisual patterns.

Ermolaeva clarifies that Eisenstein's original conception of audiovisual montage is in fact "a complex course of *interweaving* of accents of action and music, where *synchronization* is only a rare and exceptional phenomenon."[13] Just as Ermolaeva noted, this dissertation also treats sync points in animated musicals as "exceptional phenomenon," though more frequent than "rare." It is only in studying exceptional phenomenon that exceptional observations can be made.

Music theory's language and logic have largely developed into two music-focused approaches: tonal techniques borrowed from common practice and concert music, and the other based on post-tonal techniques, specifically transformation theory and pantriadicism. Scholars such as Alfred Cochran (1986), David Neumeyer (1990), Ronald Rodman (1998), Matthew Bribitzer-Stull (2015), and Tahirih Motazedian (2016) have all produced some form of large-scale, tonal, or melodic-based analysis of film music.[14] Their analyses favor music by classic film composers such as Aaron Copland and Bernard Herrmann, or composers like John Williams and Howard Shore, who follow the leitmotivic opera styles.

Large-scale tonal analysis becomes especially problematic when the New Hollywood style replaces pre-1970's Classical Hollywood music. The newer style of film scoring deemphasizes unified tonality and musical form, and musical narrative can be fragmentary or abstract without regard to create illusory musical narratives. Modern sound design, computers,

---

[13] Katya Ermolaeva, "Audiovisual Montage in Ivan the Terrible," in *Rethinking Prokofiev*, eds. Christina Guillaumier and Rita McAllister (New York City: Oxford University Press, 2020), doi:10.1093/oso/9780190670764.001.0001.

[14] Matthew Bribitzer-Stull, *Understanding the Leitmotif: From Wagner to Hollywood Film Music* (Cambridge: Cambridge University Press, 2015).

and atmospheric music allow for a more modular form of composition that problematizes tonal design, structure, and leitmotivic analyses. To address these changes, scholars like Nicholas Reyland (2016) and Steven Rahn (2019) study the lack or absence of melodic, tonal, or lyrical continuity in film music. Frank Lehman (2018) and Scott Murphy (2013) have contributed rigorous music-theoretical apparatus for analyzing non-tonal harmonies in film scores.

The academic output for sound film research in film studies and musicology exceeds that of music theory. Musicological research has tapped into audiovisual correlation, music and meaning in film narratives, cultural-historical studies, intertextual analyses, and social commentaries. Audiovisual associations support musicological commentaries with lesser focus on rigorous musical analysis. David Neumeyer calls this descriptive form of analysis "working-analyses" applicable to only a small set of film scores overseen by a single composer or director.[15] This dissertation attempts to step beyond "working-analyses" and abstract a generalized template for analyzing film rhythm based on montage (film studies), transformation (music), and rhythm and meter (music) theories. The project is angled from a rigorous music-theoretical perspective, filling in a gap identified by musicologist Robynn Stilwell in her 2002 critical literature review.[16]

[15] David Neumeyer, "Film Music Analysis and Pedagogy," in *Indiana Theory Review* 11 (1990): 5, http://www.jstor.org/stable/24045976.

[16] Robynn Stilwell, "Music in Films: A Critical Review of Literature, 1980-1996," in *the Journal of Film Music* 1, no. 1 (2009): 19, https://doi-org.uc.idm.oclc.org/10.1558/jfm.v1i1.19. Stilwell identifies a music-theoretical gap, writing "While it is true that film studies and musicology have been largely uninterested in film music—or, to be fair, ill-equipped theoretically to approach the synthetic form—the vociferous plaints of complete neglect are gradually being exposed as hyperbole."

1.2: Literature Review

Film montage theory and transformation music theory have developed sufficiently but independently of each other. Until recently, animation theories and philosophies were often scattered across film studies literature, receiving less serious academic attention than their live counterpart.[17] Film theory's predecessors in literary analysis and narratology have preceded and sometimes inspired music-theoretical studies of film music. Musicology constitutes the bulk of film music literature on an array of topics, outweighing the amount of rigorous music theoretical literature presently available. This dissertation seeks to expand the choices of available image-sound graphing by providing a framework that systematically measures audio and visual components with content determined by an eclectic constellation of theories. These theories are categorized into three broad areas: film studies related to Eisenstein's montage, theories of rhythm and meter involving projection and hypermeter, and film music studies involving transformation or pantriadic theory.

An influential leader in early Soviet cinema, Sergei Eisenstein theorizes five general methods of montage created through dialectic relations underscoring his belief that two images of similar *or* opposite meaning interact to create new meaning through intellectual association. The images, processed at certain rates and rhythms, either facilitate or obscure the immediacy of a message beyond the montage sequences. While Eisenstein is not explicit about it, this project argues that Eisenstein's prolific writings demonstrate his reliance on poietic, unnatural

[17] David Davies notes back in 2019 that "Until very recently, little if any serious philosophical attention has been paid to animated cinema." In David Davies, "Animation," in *The Palgrave Handbook of the Philosophy of Film and Motion Pictures*, eds. Carroll Noel, Laura T. Di Summa, and Shawn Loht (Cham: Springer International Publishing, 2019), 173, doi: 10.1007/978-3-030-19601-1_8. Davies attributes film's distancing from animation to an urgency to make film a "serious…potentially artistic medium" separate from the notion of photography and moving photographs.

Additionally, Paul Wells seeks to "reposition the animated film in America within a range of critical frameworks…" In Paul Wells, *Animation and America* (New Brunswick: Rutgers University Press, 2002), 1.

production-based techniques to create an organic final product that is esthesic, or reception-based. Of his contemporaries, Eisenstein's dualistic approach is strongly based on the unity of image oppositions. But as Eisenstein and his fellow director, Pudovkin, developed their theories, they moved from a more unnatural and editing-based approach to a more organicist point of view.[18] Their montage theories engaged in a cyclic interaction with American cinema, and the exchanges have embedded Soviet montage techniques and conflated montage with American continuity editing in Hollywood cinema.[19] These techniques form the basis of determining constituent components of film rhythm.

The concepts of *film rhythm* and *audiovisual synchronicity* are relatively unstandardized, being situational to the literature or original sources they appear in. "Audiovisual synchronicity" is a translation of Sergei Eisenstein's idea of vertical montage, or the intellectual information created from audio and visual interactions in a scene. Analyzing *Alexander Nevsky*'s "Battle on the Ice" cue prioritized music as the audio component, contrasting how mainstream live action films of that time were primarily vococentric.[20] While this may be an unintended product of emphasizing Prokofiev's score and involvement in the film, Eisenstein's work has since inspired a handful of audiovisual analytical frameworks.

Altman's 2013 article compiles image-sound graph frameworks and categorizes them according to the graphs' various foci: score-based analysis (Eisenstein 1942, Schaeffer 1946, Manvell and Huntley 1957), theme-based correlations (Gorbman 1987), and scoreless analyses based on timestamps, sound waves, and/or word descriptions (Altman 2000, Beck 2000).[21] Tom

[18] Vance Kepley, "Pudovkin, Socialist Realsim, and the Classical Hollywood Style," in the *Journal of Film and Video* 47, no. 4 (1995): 3-16, http://www.jstor.org/stable/20688078.
[19] David Bordwell and Kristin Thompson, *Film art: an introduction* (Boston: McGraw-Hill, 2004), 478–79.
[20] Sergei Eisenstein, *The Film Sense*, trans. Jay Leyda (New York: Harcourt, Inc., 1943), 174–216.
[21] Altman 2013, 4–20.

Schneller's modular framework for analyzing Bernard Hermann's score to *Psycho* combines both audiovisual graphs (organized by shot number) as well as musical material to identify modules and recurring stepwise dissonance resolutions.[22] These audiovisual graphs often project ideas that are locally effective, although modern Hollywood scores may require more hermeneutic stitching to make large-scale claims.

Large-scale film music analysis has primarily relied on traditional, or tonal-based, music analysis. In his 1986 dissertation, Alfred Cochran studies Aaron Copland's film music to form a framework centered on foreground, middleground, and background findings, applying traditional analytical apparatus to early film music.[23] David Neumeyer's 1990 response to Cochran's dissertation critically examines the issues and limitations of this type of film music analysis, in addition to outlining pedagogical challenges of teaching film music analysis.[24] Neumeyer identifies core issues of tonal film music analysis, including how "…one has to admit that film is an inhospitable environment for analytic methods which were designed for concert music…. serious reinterpretation is required to make these tools fit for use…"[25]

Ronald Rodman's 1998 article proposes tonal closure as a means of relating tonal design to narrative.[26] However, film soundtracks were not always congruent to the visual or plot narrative, problematizing large-scale film music analysis. Tahirih Motazedian's 2016

---

[22] Schneller, Tom. "Easy to Cut: Modular Form in the Film Scores of Bernard Herrmann," in *Journal of Film Music* 5 (2013): 127-151.

[23] Alfred Cochran, "Style, Structure, and Tonal Organization in the Early Film Scores of Aaron Copland," Ph.D. dissertation, (The Catholic University of America, 1986).

[24] Neumeyer 1990.

[25] Neumeyer 1990, 27.

[26] Ronald Rodman, ""There's No Place Like Home": Tonal Closure and Design in The Wizard of Oz," in *Indiana Theory Review* 19 (1998): 125-43. http://www.jstor.org/stable/24044541. Rodman's additional writings on tonal approaches to film music also includes a study of Herbert Stohart's music. See also, Ronald Rodman, "Tonal Design and the Aesthetic of Pastiche in Herbert Stothart's Maytime," in *Music and Cinema*, eds James Buhler, Caryl Finn, and David Neumeyer, (Hanover: Wesleyan University Press), 187–206.

Neumeyer has also produced his own tonal design analyses in film music. See also: David Neumeyer, "Tonal Design and Narrative in Film Music: Bernard Herrmann's 'A Portrait Of Hitch' and 'The Trouble With Harry,'" in *Indiana Theory Review*, 19 (1998): 87-123.

dissertation, attempts to repair this gap by distinguishing tonal design and structure to relax common practice rules on key relationships to project musical relationships on large-scale levels.[27] Such reinterpretation would be necessary according to Neumeyer's 1990 statement. Even so, what Frank Lehman considers the "score-as-a-whole" approach (opposed to "the cue") has the danger of forcing film music to conform to prescriptive theories like Schenkerian analysis or to make structural claims based on strict tonal relations.[28] Taking the cue-based approach would then require some form of surface-level (phenomenal) analysis that draws from theories of rhythm and meter as well as film rhythmic editing.

For the purposes of this dissertation, film rhythm is understood from an editing standpoint, referring to the pacing of content within a shot or the temporalities between shots. In a 2008 written symposium on editing in various media, American filmmaker Frederick Wiseman defined "internal rhythm within a sequence" and "external rhythm between the sequences" to reduce and sequence film materials into a digestible unit.[29] Bordwell & Thompson's rigorous undergraduate textbook, *Film Art*, offers a reference for analytical methodologies for studying continuity editing and the effects of internal and external film rhythm interactions.[30] Their points resonate with Karen Pearlman's three materials of film rhythm at an editor's behest, "time, energy, and movement" that push beyond notions that film rhythm is a purely phenomenological experience.[31] Therefore, the core of Wiseman's ideas, that there are external and internal rhythms

[27] Tahirih Motazedian, "To Key or Not to Key: Tonal Design in Film Music," Ph.D. dissertation, (Yale University, 2016).
[28] Frank Lehman, "Methods and Challenges of Analyzing Screen Media," in *The Routledge Companion to Screen Music and Sound*, eds. Miguel Merra, Ron Sadoff, and Ben Winters (New York: Routledge, 2017b), 497–516.
[29] Frederick Wiseman, "A Symposium on Editing: On Editing" In *The Threepenny Review*, no. 113 (2008): 16-19.
[30] David Bordwell and Kristin Thompson, *Film art: an introduction* (Boston: McGraw-Hill, 2004).
[31] Karen Pearlman, "On Rhythm in Film Editing," in *The Palgrave Handbook of the Philosophy of Film and Motion Pictures*, eds. Carroll Noel, Laura T. Di Summa, and Shawn Loht (Cham: Springer International Publishing, 2019), 143 – 63, doi: https://doi.org/10.1007/978-3-030-19601-1_7.

to film achieved through forms of editing, are contingent on the poietic and esthesic *projection* of film rhythm from an audiovisual aspect.

During the 18th century, Aristotelian concepts of time (as subdivision of discrete containers of temporal durations) switched to Newtonian time that is linear, dynamic, and individualistic (experienced differently by each person).[32] The top-down view of mensural notation from the medieval and Renaissance is replaced by more phenomenological views on how patterned rhythms create a sense of structure. In the 19th century, Moritz Hauptmann's Hegelian views of dualism contextualize his cognition-approach to rhythm and meter. He observes rhythm and meter as groups and units that create newly composite meanings.[33] This idea of oppositions that create a new whole, like Eisenstein's montage dialectic philosophies, serves as the first systematic theory of rhythm and meter. Hauptmann's theories are discussed in Christopher Hasty's 1997 book *Meter as Rhythm*.[34] Hasty's theories of rhythmic and metric *projection* expands the accent theory ideas of Hauptmann, who posits that rhythm is determined by a collective of beats. While Hauptmann himself seemed unaware of cognitive projection, Hasty systematically illustrates how relationships of one beat to another create expectations, or projections, of metric hierarchies.

Like Hasty, Danuta Mirka's (2009) and Justin London's (2012) theories also concern how listeners project time and chunks of musical information. Mirka's and London's theories reflect new ideas that people and music can shape the understanding of time, rhythm, and meter. While Mirka argues that meter is established by composers through the score, London provides a

---

[32] William E. Caplin, "Theories of musical rhythm in the eighteenth and nineteenth centuries," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen, (United Kingdom: Cambridge University Press, 2002), 657 –694. In his chapter, Caplin provides a detailed overview of the transition from 18th-century mensural systems to the modern 19th-centruy systems of rhythm and meter.
[33] Moritz Hauptmann, *The Nature of Harmony and Metre*, trans. W. E. Heathcote (London: S. Sonnenschein, 1888).
[34] Chirstopher Hasty, *Meter as Rhythm* (New York: Oxford University Press, 1997).

framework of cognitive templates and entrainment with the belief that meter is not always notated.[35] Their theories of projection allow for close study of rhythm at the local level with a cognitive bent.

Sharing concerns similar to Mirka and London, film scholars like Danijela Kulezic-Wilson and Karen Pearlman study rhythmic editing in relation to perception and cognition.[36] Both relate concepts such as time, pacing, energy, repetition, and movement to editing choices and the effects it has on the final presentation of the film. [37] Gilles Deleuze's 1983 theories of movement-image is similarly based on perception and offers a processual framework for the reception of images.[38] Beginning with a "zero-point," Deleuze suggests that viewers experience firstness, secondness, thirdness, and intermediary stages in associating a sequence of images. However, resultant analyses of either music or film rhythmic projection are potentially fragmentary, unless recurring projection schemes are reflected at larger levels, the *hypermetric* level, through studying phrase rhythm or architectonic hierarchies among rhythmic cells. Like film music studies, there are hypermeter theories based on tonal or generalized frameworks.

Following the tonal Schenkerian tradition, William Rothstein's writings on hypermeter (1989) overlaps the development of projection theories, and is defined by his concept of *phrase rhythm*, or the larger grouping of four-bar units according to some tonal or durational parameters.[39] Rothstein's works inhabits the same circle as that of Carl Schachter, Heinrich Koch, and other 18th century theorists who favored Schenkerian methodologies and 4-bar

---

[35] Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* (New York: Oxford University Press, 2012). See also Danuta Mirka, "Musical Meter between Composition and Perception," In *Metric Manipulations in Haydn and Mozart* (New York: Oxford University Press, 2009), 3–30.
[36] Pearlman, 146–49.
[37] Danijela Kulezic-Wilson, *The Musicality of Narrative Film* (Basingstoke: Palgrave Macmillan, 2015).
[38] Gilles Deleuze, *Cinema 1*: *The Movement-Image* (Minneapolis: University of Minnesota Press, 1986). His ideas are expanded in the following volume: Gilles Deleuze, *Cinema 2: The Time-Image* (Minneapolis: University of Minnesota Press, 1989).
[39] William Rothstein, *Phrase Rhythm in Tonal Music* (New York: Schirmer Books, 1989).

phrases. Any irregular phrase lengths, like 5-bar phrases requires a degree of reinterpretation or the acknowledgement of phrase expansions, or what Jarvis and Peterson term tonal or rhythmic *detour* and *reroute*.[40]

Since modern film music proved difficult for strict Schenker-based phrasal analysis, it is possible to analyze film rhythm based on architectonic concepts drawn from Fred Lerdahl and Ray Jackendoff's *A Generative Theory of Tonal* Music, a guidebook for determining grouping structures.[41] Their concerns of how expert listeners make sense of musical stimuli is well-suited for the esthesic concerns of film music analysis. Moreover, Lerdahl and Jackendoff's collaboration is underscored by a reference to metrical reductions of Schubert's *Morgengruß* by David Lewin, a figure central to the development of transformational music theory.

The analysis of *Morgengruß* synthesizes both cognition and theoretical concepts. On the one hand, Lewin invests metric awareness in the informed listener.[42] On the other, he presents a framework of metrical reductions that is architectonic and generative. However, Lewin's generation of metric elements were more concerned with transformative aspects and dynamic phenomenology, such as how the smaller units become the larger whole, thus deemphasizing the content-defining and rigorous well-formedness rules set forth by Lerdahl and Jackendoff. While Lewin's model shows how one event becomes another through process, Lerdahl and Jackendoff's writings provide a more prescriptive approach to grouping structures.

Lewin's transformative attitude is best expressed in the introduction of his 1987 publication: a line graph that describes an imaginary "line *i*" that "symbolizes a characteristic

---

[40] Brian Edward Jarvis and John Peterson, "Alternative Paths, Phrase Expansion, and the Music of Felix Mendelssohn," in *Music Theory Spectrum* 41, No. 2 (2019): 187–217, https://doi.org/10.1093/mts/mtz009.
[41] Fred Lerdahl and Ray S. Jackendoff, *A Generative Theory of Tonal Music* (Cambridge: MIT Press, 1983).
[42] David Lewin, with David Bard-Schwarz and Richard Lawrence Cohn, *David Lewin's Morgengruss: Text, Context, Commentary* (New York: Oxford University Press, 2015).

directed measurement, distance, or motion from [points] s to t.[43] I believe that Lewin's pre-transformational concept of the dynamic line *i* can apply to film studies, showing how the line directs movement from one image to another. This is the underlying philosophy of this dissertation's audiovisual analytical framework.

This concept of transformation set forth by Lewin is applied to the harmonic level by Frank Lehman and Scott Murphy. In *Hollywood Harmony*, Lehman argues that pantriadicism and Neo-Riemannian analysis, techniques of transformation theory, are effective tools for analyzing affects such as wonderment in film music, especially that of fantastical or dramatic films.[44] Instead of relying on existing Neo-Riemannian operations, Murphy develops his own harmonic labelling system, that develops forty-eight pairs of tonal-triadic progression classes (TTPC) to analyze and categorize local-level triadic progressions in film music.[45] Both Lehman's pantriadic and Murphy's harmonic methodologies provide concrete labels for the musical aspects of audiovisual analysis. They are also useful for handling music that follows Nicholas Reyland's "two dominant strategies of musical scoring for affective intensities: 'corporate classicism' and 'the metaphysical style'" that respectively differentiates classic and new Hollywood scoring.[46] According to Reyland, the "metaphysical style" uses traditional formal continuity strategies,

---

[43] David Lewin, *Generalized Musical Intervals and Transformations* (New Haven: Yale University Press 1987), xxix.
[44] Frank Lehman, *Hollywood Harmony: Musical Wonder and Sound of the Cinema* (New York: Oxford University Press, 2018).
[45] Scott Murphy, "Transformational Theory and the Analysis of Film Music," in *The Oxford Handbook of Film Music Studies*, ed. David Neumeyer (New York: Oxford University Press, 2014), 471–499. Although transformation theory is the focus of this dissertation, Murphy also authored an article regarding the role of major tritone progressions and the role triadic progressions play in crafting science fiction sound worlds. His MTTP system can be seen here: Scott murphy, "The Major Tritone Progression in Recent Hollywood Science Fiction Films," *Music Theory Online* 12/2 (2006).
[46] James Buhler, *Theories of the Soundtrack* (New York: Oxford University Press, 2019). For primary source, see Nicholas Reyland, "Corporate Classicism and the Metaphysical Style: Affects, Effects, and Contexts of Two Recent Trends in Screen Scoring," in *Music, Sound, and the Moving Image* 9, no. 2 (2016): 115-130.

while "corporate classicism" typical to modern films is atmospheric and immediately affective without necessitating melodic, tonal, or lyrical continuity.[47]

One danger to larger-scale score and local-level cue analyses is the potential to abstractly analyze music without regard to immediate onscreen visuals, given the theoretical frameworks' origins in absolute music. For the matter, Stilwell's transmission of Martin Marks' 1979 criticisms on musicological research remarks how "…it is truly astonishing how many studies of the music tend to ignore completely what is happening on the screen."[48] Responding to this issue is Lea Jacobs' *Film Rhythm After Sound* that analyzes film and music rhythm on the same level, providing a combined analytical framework to study syncopated sound in film and animation.[49] Using the film score and transcriptions as the basis, Jacobs highlights musical rhythm and meter, constrained by bar lines and time signatures, as the optimal reference for measuring synchronicity. This dissertation seeks to rigorously develop ideas like Jacobs' and succeeds the lineage of image-sound graphing methodologies first inspired by Eisenstein's audiovisual score.

1.3: Key Influences

Eisenstein's audiovisual theories serve as a platform for my film rhythm-oriented framework I call *sync point analysis*. In the following chapters, I will analyze for my definition of *audiovisual rhythm*, the synchronous rhythmic interaction between image and sound, instead of film rhythm, which can exist independently of sound. Connecting disparate topics, my methodology implies that Eisenstein's Soviet montage theories from the 1940s are still relevant

[47] Ibid.
[48] Stilwell, 20.
[49] Jacobs, 74–83.

today, and that animation, which explains rhythm and flow within a shot, can fill in gaps of original film rhythm analyses that focus on relations in *between* shots and not movement *within* shots. It ultimately implies that, as music can form an illusory connective narrative throughout a sequence of disconnected images, music theory can connect different subfields of film studies.

My association of Eisenstein's theories to American and Japanese-focused animation is not unwarranted. As noted by Esther Leslie, Eisenstein was impressed of and wrote about Disney's film practices.[50] Although he was initially interested in Disney's design for synchronized sound, his interest eventually spread to the visual aspects such as motion and caricature.[51] In the very title of his translated article, "An Unholy Alliance of Eisenstein and Disney," Eiji Ōtsuka draws connections among Russian avant-garde influences, the "Disneyfication" of manga, and the increased access to American cartoons in the 1930s as key factors in the shaping of modern-day Japanese animations.[52] He also asserts that "An aesthetic unification of Eisenstein and Disney under conditions of fascism is the origin of the Japanese manga and animation that everyone today associates either with Japanese traditions or with postmodernism."[53] With these connections, I seek to apply an interdisciplinary and musical layer of understanding to the stylistic practices of Eisenstein, Disney, and Japanese *anime* through technical means, and craft my framework in a way that it can accommodate all three schools of filmmaking.

There are existing theories of sound graphing and audiovisual analysis, and my framework most combines methodologies of Sergei Eisenstein's and Manvell & Huntley's visual

---

[50] Leslie, Esther, *Hollywood Flatlands: Animation, Critical Theory, and the Avant-Garde* (London: Verso, 2002).
[51] Ibid., 244–45. Eisenstein's writings are also collected in *On Disney*, an edited compilation. See also: Sergei Eisenstein, *On Disney*, trans. Jay Leyda (London: Methuen, 1988).
[52] Eiji Ōtsuka, "An Unholy Alliance of Eisenstein and Disney: The Fascist Origins of Otaku Culture," trans. Thomas Lamarre, in *Mechademia* 8 (2013): 251-277, doi:10.1353/mec.2013.0002.
[53] Ibid., 252.

graphs.[54] In fact, it attempts to revise and update some of Eisenstein's precepts on audiovisual counterpoint (also known as vertical montage). However, my sync point framework generalizes Eisenstein's and Manvell & Huntley's more filmic focus and adds music-theoretical rigor beyond Eisenstein's and Manvell & Huntley's original graphs. This dissertation's proposed framework is also unique in relating transformation music theory to both Sergei Eisenstein's montage theories for pacing between shots and animation theories to in-shot movement. At the time of this writing, I am unaware of any English-language methodologies that combines theories of rhythm and meter, Lewin's transformational attitude, Soviet montage theory, and animation rhythm and flow into one generalized apparatus.

My framework proceeds with several generalizations that set limitations for the framework. First, my focus on film rhythm ignores non-musical audio such as dialogue, voice-over narration, and sound effects. The audio component is strictly based on film music that comes ridden with its own meanings and implications. This is a necessary simplification to focus analysis on the inner workings of music in relation to image.

Second, my analysis generalizes the source of music. There will be no differentiation between diegetic and non-diegetic music, two concepts that have flourished in film music studies from the onset of sound film.[55] This generalization is made in favor of studying the immediate effects of synchronous audiovisual rhythm, especially when diegesis is not immediately

---

[54] Altman 2013, 4–10. For the original source, see also Roger Manvell and John Huntley, *The Technique of Film Music* (New York: Hastings House, 1957).

[55] David Neumeyer explains diegetic music as "Sound belonging to (anchored in) the physical world depicted in the narrative [film]." Brackets, mine. He contrasts this with nondiegetic music, as "Sound belonging to the level of narration or the narrator." In other words, diegetic music is what characters can hear. Nondiegetic are what the characters cannot hear, but the narrator or filmgoer can. Neumeyer considers diegetic/nondiegetic as one of five binaries of film music functions. For more, see Table 2.2 in David Neumeyer, *Meaning and Interpretation of Music in Cinema* (Bloomington: Indiana University Press, 2015), 63.

recognized by a film viewer. This also means that synchronous sound is generalized with no special attention to concepts of mimesis, fidelity, and sound source.

Third, the rhythmic surface of music is generalized. Using graphic protonotation such as nodes and lines will generalize visual events as happening on a structural beat of the indicated measure and time signature, disregarding the lowest-level durations. This generalization is made to avoid overanalyzing local level events occurring within a single beat.

Fourth, external techniques that connect between-shots are restricted to labels from Eisenstein's montage theories, specifically rhythmic and metric montage. These are the most local-level montage types and are emphasized as immediate indicators of film rhythm. The internal rhythm of the shots is limited to two factors: key-poses and character movement. Other internal features such as a shot's mise-en-scene, shot type (long shot, medium close-up, wide shot, etc.), and lightning are de-emphasized in favor of studying rhythm. All case studies in this dissertation point towards a relationship between the rate and frequency of accents on the musical surface and the frequency of visual film rhythm techniques.

While he did not write about a transformational attitude in the manner of David Lewin, Eisenstein's theories of audiovisual counterpoint apply the idea of transformation to the audiovisual whole created by audiovisual rhythm. James Monaco writes that "[To Eisenstein,] The process of film (like the process of theory) was far more important than its end, and the filmmaker and observer were engaged in it dynamically."[56] The process and engagement are the connection between shots A and B, like Lewin's dynamic line *i*. Audiovisual interaction possibilities that can link shots A and B include acceleration of film rhythmic cuts, camera

[56] James Monaco, "Film Theory: Form and Function," in *How to Read a Film* (New York: Oxford University Press, 2000), 403. Brackets, mine.

motion, musical harmonies, hypermetric musical events, or a combination of said techniques. I believe that analyzing these musical sequences, figuratively the dynamic process represented by line *i* connecting music and image, is motivated by a sense of motion. This concept of motion is dynamized by editing techniques reliant on executing *tension* through synchronous ambiguity, *energy* through moving forward in ambiguity, and *resolution* in synchronous restoration at the downbeat (hypermetric montage) or with illusive stability at audiovisual cadences (increased synchronized cutting and harmonic activity). Such relations resonate with Karen Pearlman's "time, energy, and movement," that contribute to building visual tension and release.[57]

My application of transformational music theory to film music intersects with the research of theorists such as James Buhler, Frank Lehman, and Scott Murphy. Although music theory has a smaller output of film rhythm studies, general analytical frameworks are viable to understanding surface- to deep-level audiovisual synchronicity. I rely on transformational harmonic analyses theorized by these scholars but apply it to the dynamic filmic surface happening *in time*, as opposed to studying transformations solely in abstract space. This aligns with the projection theory of Hasty but applies to the dynamic viewpoints of Mirka and London. Lerdahl and Jackendoff's architectonic approaches to rhythm will be combined with William Rothstein's hypermetric theories to make larger claims about modules of recurring projection schemes that emerge from analysis. The resultant idea of applying transformation theory to film and music rhythm taps into relevant scholarship on rhythmic film editing, as laid out in the Literature Review section. Eisenstein's montage theories, supported by analyses and

[57] Pearlman, 148.

philosophies of rhythmic editing, will serve as the content for analyzing rhythm between shot—or the external rhythm defined by Wiseman.

Yet, the constant use of camera or character motion may challenge a straightforward editing-only approach to analysis. The rate of rhythmic cutting may be less straightforward in a corpus of diverse film genres and styles. There may be no consistent gap of time between one cut or the other, and camera motion may substitute areas where rhythmic cutting is. Therefore, for internal rhythmic content, I use animation theory to analyze in-shot key-poses through perception and projection of manufactured *motion* from key-poses A to B. Animation theory's concepts of accent, anticipation, and analogy of motion are also analogous to Lewin's dynamic line *i*. In their 1981 book, Disney animators Frank Thomas and Ollie Johnston delineates twelve timeless principles of animation.[58] Among the principles, "anticipation," "straight ahead" and "pose-to-pose," "follow through," "slow out," and "slow-in" are various rhythmic techniques that animators use to move and transform one character pose to another. Though separate from the soundtrack, these techniques can be exploited at the audiovisual level.

Music theory and transformational harmony can stitch together the external betweenness of film montage and the internalizations of animation theories. Film's constant visual invitations for image-projection is like music's constant invitation for audio-projection, as demonstrated by the writings of Christopher Hasty. I believe that transformational motions of musical, filmic, and amination levels are additionally isomorphic to Hasty's concept of rhythmic-metric projection.[59] Treating audiovisual film as an indeterminate experience still requires a level of temporal measurement. Therefore, the inherent measurability of musical lengths in terms of measure

[58] Frank Thomas and Ollie Johnston, *The Illusion of Life: Disney Animation* (New York: Hyperion, 1995).
[59] Christopher Hasty, *Meter as Rhythm* (New York: Oxford University Press, 1997).

numbers, note values, time signatures, and tempo markings allows music to serve as a measurable basis for measuring between- and in-shot motions. Music-based temporal measurements allow for optimal alignment of sync point and shot references. Taking a sound film playback as the product, my framework develops a system that examines the occurrences of synchronous points in experienced *technical* time spans.

1.4: Chapter Previews

This dissertation can be generalized into two broad sections that present the sync point framework in a spiraled methodology, from basic concepts to advanced applications. Part I serves as a taxonomical toolkit where I develop a framework for *sync point* analysis that (1) differentiates synchronous from nonsynchronous audio-image relations and (2) assigns labels to describe sync point functions. I introduce the placeholder and plugin formula n[x]-[y] as the functional label for sync points. Chapters 2–4 help define different parts of the formula whereby *n* stands for rate of rhythmic occurrence, container [x] stands for music accent type, and container [y] stands for visual accent type.

In Chapter 2, I introduce the idea of *rhythmic sync* to describe any form of synchronicity that is nonperiodic and unpredictable. These "R-type" sync points are then described by the type of musical accent [x] and visual accent [y] synchronized at a precise, given moment. Musical accents [x] may correlate to a sonic feature on the musical surface, phrase, or durational value. The visual accents [y] may synchronize to the music feature with an image that is isomorphic (duration related), iconic (abstractly related), or a combination of both. The taxonomy introduced in this chapter serves as the building blocks for the rest of the dissertation. Said taxonomy is

drawn from and analytically applied to case studies that primarily draw from Disney musical series such as *Fantasia 2000*, *Hercules* (1997), *Mulan* (1998), and *Frozen* (2013). Analysis is also extended to Japanese *anime* such as the *Revue Starlight* (2018) series and the *Hatsune Miku: Colorful Stage!* (2020) multimedia series. A focal analysis compares two animated adaptations of Camille Saint-Saens's "Finale" from *Carnival of the Animals* to demonstrate how labelling sync points provides a narrative play-by-play that compares Disney's fluid mickey-moused narrative in *Flamingos* to the more fragmented and slapstick narrative of Warner Brothers' *Bugs and Daffy's Carnival of the Animals*.

Chapter 3 introduces the *audiovisual* [AVC] cadence and *vococentric* [VCO] shot change as more specific functional subsets of the rhythmic "R-sync" types defined in Chapter 2. Case studies demonstrate how rhythmic acceleration and groupings of sync points work together with music to emphasize important climactic moments in an animated musical. Many of these moments involve the cadential passage, and I introduce the term *audiovisual cadences* to describe aurally and visually emphatic moments of music-formal closures. As a cadential event, [AVC] and all its forms create synesthetic formal closure by highlighting cadential passages and the cadence proper with shot changes, camera motion, and/or character motion. Examples such as "Honor to Us All" from *Mulan* (1998) serve demonstrate how camera motion can substitute shot changes and accommodate evolving technological advances in the field. At the same time, the substitution reinforces shot changes as the foundational technique of film rhythmic practices.

The vococentric [VCO] visual accent is also introduced in Chapter 3 to accommodate visual funneling and attention to sonic fidelity in multi-member singing ensembles. Examples of "Zero to Hero" (*Hercules*, 1999), "Cinema" (*Colorful Stage!*, 2020), and "Love is an Open Door" (*Frozen*, 2013) demonstrate [VCO] type sync for songs with lyrics. The focal case study

of "Turkey in the Straw" from *Steamboat Willie* (1928) then takes the [VCO] principles established in the prior examples and applies it to an instrumental passage that is "voiceless" and specific to the different timbres present in the scene. Even so, the sync point framework can map the same underlying principles of sonic fidelity to that of cartoon characters, whose "voiceless" voices are grounded in different instrumental timbres.

Chapter 4 expands on the n[x]-[y] function label by adding the possibility of *metric sync* that is periodic and predictable. I demonstrate how recurring sync points that are equidistant and equal in duration form organizing structures of two-, four-, and multi-bar phrases. These structures form sync point groups at the local level and open pathways towards making larger-scale observations regarding film rhythmic profiles and musical form. Although based on musical terms, the metric and hypermetric sync serve as functional labels for audiovisual rhythms that emerge from synchronicities and correspond to, while being independent of, musical hypermeter. The terms also largely rely on Eisenstein's definition of "metric" as being shots that are equi-durational and equidistant from one another.

(Hyper)metric sync is determined by a tripartite process that mirrors the rhythmic projection theories of Hasty, Mirka, and London, all while applying the well-formedness concepts of Lerdahl and Jackendoff. Chapter 4's case studies adopt an energistic approach to demonstrating the organizational and narrative-framing function of (hyper)metric sync. In a spiraled fashion, *Flamingos* from *Fantasia 2000*, "Cinema" from *Hatsune Miku: Colorful Stage!*, and "Turkey in the Straw" from *Steamboat Willie* are revisited to demonstrate how R-sync can *elevate* to (hyper)metric sync under certain conditions. Additional examples from *Steamboat Willie* and *Mulan* show how (hyper)metric sync can occur anywhere, from formal organization to the local cadential [AVC]. The focal case study touches on the "Battle on the Ice" cue from

Alexander Nevsky (1938) that underlie Eisenstein's theories of *audiovisual* synchronicity, and by extension, all audiovisual frameworks that succeeded his writings. The case study demonstrates that analyzing for (hyper)metric sync helps outline a loose binary form between the metric and non-metric sync points, as well as the underlying Neo-Riemannian harmonic progressions. All these independent theories and analytical framework are united under the banner of sync point analysis, demonstrating the inclusiveness and flexibility of the apparatus.

Part II presents advanced analytical applications of the sync point framework. Once the n[x]-[y] opens pathways to analysis, larger-scale observations can be made of form and cadential action. Chapter 5 illustrates how sync point types can serve as building blocks for *sync trains*, or a visual map of sync points and the in-between actions that connect them. The sync trains not only provide visual summaries for the emergent rhythmic narratives of sync points in film and music videos.

Chapter 6 pushes analysis further by using sync trains as a method to analyze mobile rhythm games and open pathways to understanding sync points in a haptic (touch-based) dimension. The case study of "Close to Gray" from the *Hatsune Miku: Colorful Stage!* series pushes the boundaries of my framework to accommodate mixed media animations and demonstrates how my framework is future-oriented and adaptable to technological advancements. In Chapter 6, I apply theories of rhythm and meter to beatmap gameplay, explore haptic (touch-based) synchronicities between player and media, and form a skeleton for trilateral sync. The addition of a third rhythmic stream, the beatmap, complicates sync point analysis with a three-way synchronicity between music, visuals, and player-driven gameplay. And yet, the sync point analytical framework is able to resolve analytical conflicts through principles developed in Part I of this dissertation.

All chapters of Parts I and II demonstrate the robust analytical capabilities of relating musical harmony to rhythmic film editing as well as the benefit for contributing interdisciplinary systems to a body of existing visual analysis methodologies. My case studies aspire to show that sync point analysis can tap into any form of animated multimedia, including the relatively untouched repertoire of Japanese anime to define over-synchronized audiovisual elements (i.e., "mickey-mousing") as more than plot accessories. In assigning definitions to the n[x]-[y] plugin formula, I develop my framework in hopes to demonstrate how studying points of mickey-mousing, mimicry, or mirroring is not a mere plot device but a wellspring of emergent meanings that actively defines and redefine rhythmic aspects and receptions of the filmic experience.

Part I: Taxonomical Toolkit

Chapter 2

Rhythmic Sync Points

2.1: Background

In his 1992 chapter on cartoon music, Scott Curtis critiques film scholarship's "feature-centric" model and its taxonomical inadequacy for describing certain aspects of animation. Curtis identifies three schemata that limit analyzing audiovisuality in cartoon music: "the image/sound hierarchy, the separation of the sound track into dialogue/music/effect, and the 'diegetic/nondiegetic' distinctions."[60] This chapter responds to Curtis by presenting sync points as purposeful analytical foci and establishing a taxonomical framework for determining sync point rhythmic *functions*. Through categorizing musical and visual accent techniques found in animated musical sequences, I will demonstrate how music theory principles help identify *what* is being synchronized while film-visual techniques inform *how* the music is "mickey-moused," or closely mirrored. In this presentation, I make the claim that mickey-mousing results from a high occurrence, or frequency, of functional sync points that bears significance in defining

[60] Scott Curtis, "The Sound of the Early Warner Bros. Cartoons," in *Sound Theory, Sound Practice*, ed. Rick Altman (New York: Routledge, 1992), 192–201. Curtis makes a strong case for why traditional schemes for analyzing film is complicated by production technicalities of animation.

In his chapter, Curtis explains the sound/image hierarchy whereby the sound (i.e., music) is often recorded after the image component is completed, takes secondary role to the image, and is therefore "supplemental" (194). However, production practices in animation complicates his straightforward relationship when music and sound precede the image, which may even be timed to a click track. Likewise, early animation productions dealt with technological changes that resulted in mixed methods of recording. The use of the Vitaphone (a sound-on-disc system as opposed to sound on film strips) recorded all sounds together, thus making sound a composite package instead of separate sound streams: Curtis writes "However, the Vitaphone recording system in use until 1933 required all elements of the sound track be recorded at once" (197). Finally, animation can creatively circumnavigate fidelity to sound sources and can feature caricature expressions supported by sound effects, background music, and even singing voices that are not necessarily *of* the animated character's lived world. This is why Curtis also argues that the "diegetic/non-diegetic" scheme is ineffective in dealing with the sonic profiles of animation.

functional audiovisual interactions.[61] The sync points may even reveal stylistic choices of different animation studios. Though the term mickey-mousing may carry negative connotations,[62] I seek to push mickey-mousing beyond its pejorative status as pure mimicry or plot accessory and present its essence, the existence of audiovisual synchronicity, as a dialectic tool that actively creates emergent meanings in an animated, synesthetic, and multimodal experience.

Different from traditional live-action film analysis, studying animation is productive when image and sound are examined non-hierarchically, on equal terms, and understood as a fluctuating cycle of influence. Image and sound interact with each other in time, and their simultaneous and reciprocal influences can quickly interchange the salience of audio and image, insomuch that assigning hierarchy to audio and image becomes cumbersome. Animation's ability to deemphasize the necessity for dialogue contrasts with David Neumeyer's transmission of Michel Chion's study on live action sound film as "voco- or verbocentric," or the prioritization of voice and dialogue above all other sounds in the soundtrack.[63] While cartoons can be

---

[61] Lea Jacobs, *Film Rhythm after Sound: Technology, Music, and Performance* (Oakland: University of California Press, 2015), 58. According to Jacobs, "Mickey-mousing" is a process and term that "encompasses a number of different aspects of the relationship between music and action, and music and other sounds. Most important for my purposes is the idea of a tight synchronization between movement and/or cutting and the beat. But, it is also used to refer to the musical imitation of physical movement…. [and] is also sometimes applied more generally to any tight integration of music and sound effects…"

[62] Curtis, 201. Here, Curtis notes that the term "mickey-mousing" often carries a "pejorative" status.

[63] David Neumeyer, *Meaning and Interpretation of Music in Cinema* (Bloomington: Indiana University Press, 2015), 12. Neumeyer's transmits Chion's sound track hierarchy where the voice carries foremost priority. Neumeyer cites Chion makes remarks such as "'The voice hierarchizes everything around it,'" or that human beings in general focus on the voice because it is a "habitual behavior."

Neumeyer furthers that "Claiming film music for a discipline by constructing interpretations of implicit meanings grounded in the idea that music is 'equal' to the image, or 'agential,' will only work if one also acknowledges the limitations of the contexts in which such claims can be made, or, to put it another way, if one acknowledges the limitations imposed by an inevitably distorted mode of viewing and hearing a film" (13). This chapter acknowledges Neumeyer's viewpoint and seeks to investigate the limited contexts such as animated musical numbers or adaptations of orchestral works that are not primarily vococentric. These limited contexts focus on standalone musical numbers within larger Disney films or sequences that support Paul Wells' idea that animation itself can be an abstract or distorted mode of "viewing," or directing "viewing" of a film. Through the case studies in this chapter, I believe that animation as an art form presents and invites what Neumeyer considers "an inevitably

primarily "voco- or verbocentric," the dissertation's framework is designed for case studies that are standalone performances, songs, and music videos where the speaking voice (including dialogue) is absent, blurring distinctions of the diegetic and nondiegetic. Holding onto Curtis's non-hierarchical philosophy avoids the chicken-or-egg question of whether music or image came first during production, especially when analyzing animations that do not always provide their production contexts and processes.

In this dissertation, the term "synchronicity" refers to any point in an animated musical sequence where image and sound closely mirror each other in a rhythmic manner of fluctuating periodicities. In every synchronized animation sequence, there are two immediately salient rhythmic streams: the rate of visual accentuation and the musical surface rhythms, marked by melodic and phenomenal accents. In Figure 2.1 from "Flamingos" (Disney's *Fantasia 2000*),[64] the rhythmic line represents the "visual" stream, while the piano reduction accounts for the "audio," or the musical reduction of the introductory passage from Camille Saint-Saëns' "Finale" from *Carnival of the Animals*. Both streams are similar in that they accelerate in rhythm and duration, be it the transition from tied whole-note values to half notes in the "visual" stream or

distorted mode of viewing and hearing a film." For more on Wells, see: Paul Wells, *Understanding Animation* (New York: Routledge, 1998).

[64] This analysis is based on the Disney *Fantasia 2000* animated short. The original video can be found on DVD, which is the reference for this analysis. See: Walt Disney Home Entertainment, "Flamingos," in *Fantasia 2000*, directed by Eric Goldberg (Burbank: Buena Vista Home Entertainment, 2000), DVD, 74 min. The video is also available on Disney+ at the link: https://www.disneyplus.com/movies/fantasia-2000/5DG1yXGiy3m5 (last accessed June 2024).

Measure numbers used in analysis, going forward, refers to the actual score of the "Finale" from *Carnival of the Animals* by Camille Saint-Saens. While the author transcriptions and reductions try to mirror that of the *Fantasia* version, two scores were additionally referenced to ensure accuracy of notation. The scores referenced are as follows: Camille Saint-Saëns, *Le Carnaval des Animaux*, ed. Auguste Durand and Fils (Paris: Durand, 1922), accessed June 2024, https://vmirror.imslp.org/files/imglnks/usimg/0/06/IMSLP576538-PMLP6099-Le_carnaval_des_animaux_-...-Saint-Sae-ns.pdf.

There is also a piano, four hands version. See also: Camille Saint-Saëns, *Le Carnaval des Animaux*, arranged for piano, 4 hands by Lucien Garban (Paris: Durand & Cie., 1922, repr. Boca Raton: Masters Music, n.d. 1998), accessed June 2024, https://ks15.imslp.org/files/imglnks/usimg/1/10/IMSLP13513-Saint-Saens_Carnaval_des_Animaux_piano_4_hands.pdf.

the addition of the orchestra's motivic patterns in the "audio" stream. Neither stream implies musical or phrasal function in terms of traditional music-theoretical principles of phrase structure. How, then, would this passage be considered *functional*, in terms of audiovisual relations?

In this study, *function* is defined by a set of conditions and expectations formed by recurrence or non-recurrence of emergent meanings. The *meanings* are dialectic and emerge from audiovisual interactions like that in Figure 2.1. Although the two streams are seemingly non-functional on paper, the visual accents highlight specific musical aspects such as alignment with the downbeat, duration, and rhythm. In Figure 2.1, arrows indicate *sync points* and detail the interacting audiovisual accents in the excerpt. The arrows relate specific parts of the music to the rhythmic rate of occurrence on the "visual" line. The transparent gray numbers tally the synchronous instances. The sun rays in mm. 1–4 appear and expand with the piano tremolos, conditioning meaning ("sun rays") to the piano's timbre and speed. A third iteration of "sun rays" is expected in mm. 5–6, but this expectation is reconditioned when the flamingos begin to move at sync point 4 (m. 5, beat 3). Partway into the "sun rays" tremolos, the visual accents abandon their first audiovisual association and create new meaning ("waking up") by accenting the orchestral G-C#-D motives. While the "sun rays" and "waking up" can simply be regarded as mickey-mousing, I posit that these interactions are *functional*. The visuals add meaning to the action, and the musical tremolos and motives add depth to the visual motion. Both audio and visual components enable a cyclic creation of associative meaning for sound and image.

Sync points are the *conditions* for function, as *expectations* formed by the conditions create emergent meanings. Because the sound experience and representation of sounds are varied and individualistic to listeners, sync points provide a technical form of interaction that create

meaning when they match visual to musical actions and events. In Figure 2.1, sync points indicated by numbers 1, 2, and 3 associate visual information ("sun rays") with strong beat tremolos and the piano timbre. The quickening sync points 4–7 visually accent the flamingos to emphasize the orchestral timbre and its syncopated nature; the whimsical melodic tidbits offset the phrase and reorient listeners to motion beginning on the third beat of each measure. Additionally, this introduction foreshadows the audiovisual tug-of-war unfolding between a unique yo-yo flamingo, represented by piano, and the common flamingos, represented by the orchestra. This contention between timbres is set up from the introduction, with common flamingos sonically conditioned to the orchestra while the piano is conditioned to the "other," or what is not part of the common flamingos. This timbral conditioning sets up expectations and audiovisual associations that facilitate the narrative of "Flamingos" from *Fantasia 2000*.

The preliminary analysis of "Flamingos" illustrates only one form of functional audiovisual synchronicity. There are innumerable ways that two non-functional rhythmic streams can work together to create emergent functional meanings through sync points that highlight clearly functional musical passages. Figure 2.2 presents the climactic ending of "Zero to Hero" from *Hercules* (1997).[65] The passage is clearly a musical cadence, with an ascending melodic phrase that formally ends the song. The music itself carries a set of conditions and expectations for a final cadence that is further emphasized by visual accentuation. The abundance of sync points, albeit clear mickey-mousing, is functional in visually emphasizing every beat of the climactic ending. The small sideway arrows indicate the use of character motion to further propel the sync point's visual markedness. Theoretically, the same excerpt could feature a single frame with Hercules sliding through a Grecian setting instead of the rhythmic shot changes. However,

[65] For media referenced for this analysis, see: Disney. "Hercules | Zero to Hero | Lyric Video | Disney Sing Along," YouTube video, 2:16, https://youtu.be/yOL-EJZjmp0?si=5zaxHmNKM-mDtPQT.

the synchronization of visual accents to each rhythmic beat in the music creates emergent meaning, conditioning the function of shot changes to the musical ascent. In what follows, I will construct a framework to examine such rhythmic audiovisual synchronizations and delineate a set of functions created by non-periodic sync points.[66]

2.2: Terminology and Definitions

My theoretical framework treats films as a hypertext where audiovisual sync creates a wellspring of emergent meanings. The concept of emergent meanings in this presentation combines Sergei Eisenstein's dialecticism with Karen Collins' application of "emergent meanings" to ludomusicological concepts and defines how media audiences form meaning through the medial reception of films.[67] The basis of my terminology is inspired by Sergei Eisenstein's writings in *Film Form* and *Film Sense* and adapts his audiovisual concepts to measure rhythmic associations. The resultant rates of musical (audio) and image (visual) changes guide interaction with a film product. To develop my framework, I combine film theories by Eisenstein with theoretical frameworks by William Rothstein (phrase rhythm) and Fred Lerdahl & Ray Jackendoff (grouping and metric structures) to create sync point labels akin to music theory's Roman Numeral analysis.

---

[66] My analytical framework examines rhythmic sync points first, prioritizing non-periodic occurrences and expectation as the building blocks of larger-level sync point relationships. One such "larger" relation is the periodic, or *metric*, sync point that will be covered and developed in the next chapter. This framework is therefore essentially bottom-up, meaning larger relations are formed by large-scale periodic recurrences of otherwise non-periodic sync points.

[67] Karen Collins, "Interacting with Sound," in *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games* (Cambridge Massachusetts: MIT Press, 2013), 27. Collins attributes the idea of *emergent meanings* to preceding scholars such as C. Wahn Park, Walter Murch, and Michel Chion—all concerned with forms of film and media studies. Collins' contribution is the application of emergent meanings to the field of ludomusicology and ludic interactive play, resulting in the more active and energetic approach that my sync point analysis takes after.

This dissertation presents functional sync points as key elements that pinpoint *how* audiovisual interactions create meaning. While sync points can simply be described as a product of interacting rhythmic streams, knowing the interacting types of visual and musical accents adds depth to analyzing audiovisual narratives. This is especially important when animation blurs the lines between the "soundtrack" and the "sound track." On the one hand, Tahirih Motazedian's research clarifies that the soundtrack refers to "a commercially released music album associated with a film."[68] On the other hand, the idea of "sound track" (and not soundtrack) is described on separate occasions by Curtis and Motazedian as an audio component made of dialogue, music, and sound effects.[69] Therefore, the sound track presents a three-point collection of categories that separates components in a film's composite audio stream.

Animation, and especially Disney musicals, can feature dialogue through song, the orchestrated music, and sound effects written into the score. For example, Figure 2.3a from "Flamingos" excerpts the percussion and orchestral clap that aurally cue the birds' collision with a tree branch. Likewise, Figure 2.3b features orchestral strings that stand in for stomping sounds. This evidence of closely intertwined musical interaction is why Curtis suggests that the "separation of the sound track into dialogue/music/effect" is unbeneficial to analyzing cartoon scores, and a reason that I will propose a more specific typology for analyzing sound-synchronized animation.[70] The terminology serves as local-level analytical labels akin to the idea of Roman Numeral analysis, whereby functions can be quickly displayed and inferred through a simple labelling system.

[68] Tahirih Motazedian, "To Key or Not to Key: Tonal Design in Film Music," Ph.D. diss., (Yale university, 2016).
[69] Ibid. Also see: Curtis, 192 and 199.
[70] Curtis, 192.

My system labels sync point functions in a formulaic manner. Displayed in Figure 2.4, the elements in the plugin formula n[x]-[y] are placeholders for categories that describe a sync point. The letter *n* stands for the *rate* of occurrence. In this chapter, *n* is a placeholder for "rhythmic," or the nonperiodic occurrences of sync points. Category [x] is a placeholder for the type of musical accent synchronized to visual action, while category [y] describes the visual accent technique that emphasizes musical feature [x]. Visual techniques include shot changes, the initiation of character motion, or the initiation of camera motion. This section provides a walkthrough of each category with examples drawn from animated musicals.

Category *n* represents the rhythmic aspect of the sync point. The first type is *rhythmic*, labelled R[x]-[y], and adapts Eisenstein's definition of rhythmic montage to describe sync points that are immediate, non-periodic, and have unpredictable rates of frequency. Eisenstein's original writings in *Film Form* explain that rhythmic montage "…gives way to a flexible relationship of the *actual* lengths [of each filmic "shot"]."[71] The flexibility of non-adherence to more mathematically measured lengths defines the non-periodicity of rhythmic sync. Therefore, visual factors of the scene would emphasize music that is non-functional or built on very local-level harmonic rhythm and phrase-structural syntax. Rhythmic "R" status is assigned when a collection of sync points does not share a durational, periodic regularity among the parts. In other words, each R[x]-[y] must be nonperiodic and unpredictable. Should they become predictable over a span of time, they will transform into a different sync type to be covered in Chapter 4.

---

[71] Eisenstein, *Film Form*, 74. In his writings, Eisenstein wrote "Abstract determination of the piece-lengths gives wat to a flexible relationship of the *actual* lengths. Here the actual length does not coincide with the mathematically determined length of the piece according to a metric formula."

Eisenstein's original definition of rhythmic montage has to do with the *content* of each film shot. He writes that "in determining the lengths of the pieces, the content within the frame is a factor possessing equal rights to consideration" (p. 73) and that the shots' "practical length derives from the specifics of the piece…" (p. 74). My analytical framework departs from Eisenstein in that my analyses *do not* forefront the contents of each shot in determining "R" status.

Introduced in Table 2.1 are definitions for Rhythmic sync types for all possible musical, visual, and combination of functions created by audiovisual synchronicities. The simultaneous occurrence of both audio and visual streams creates an interaction defined as:

R • [music accent type]-[visual accent type], where *R* = "rhythmic sync point"

With this labelling system, I indicate that R is a rhythmic sync point occurring with uneven frequency.[72] As stated before, rhythmic sync points are categorized based on their uneven frequency (or unpredictability) of occurrence, and non-functional or local-level-only harmonic rhythm and phrase-structural syntax.[73] Plugging in the rhythmic rate "R" formulates the following technical definition:

R[x]-[y] = *R • [music accent type]-[visual accent type]*

R[x]-[y] represents a basic rhythmic sync point with unspecified [x] and [y] values. Therefore, R[x]-[y] stands in for a total entity, which is a sync point that is defined by "R," the unpredictable simultaneities of [music accent] and [visual accent] types.

Music accents will be examined first, by order of the left-to-right orthography of the sync point label. For this purpose, the following section defines the components of R[x]-[y], specifically by defining the *R • [music accent type]* component of the equation. The musical content is categorized by phenomenal accents (pitch, rhythm, timbre, and other musical

[72] This definition is based on Sergei Eisenstein's idea of rhythmic montage as non-periodic occurrences of visual shot changes that do not provide a sense of temporal regularity.

[73] "Non-functional or local-level-only harmonic rhythm and phrase-structural syntax" means that the harmonies or phrases synced to sound and image must either be (1) nonfunctional in terms of music-theoretical conventions or a small excerpt of a functional progression/phrase or (2) functional at a local level without projecting hypermetric or formal implications. The latter concept will be discussed in Chapter 4.

parameters) or phrase structural entities such as motive, sentence, antecedent, and consequent. However, animation's mickey-mousing may stress any one of three general musical features:

1. Sonic (a sound event)
2. Phrasal (mirroring a phrase or phrase fragment)
3. Duration (lasting as long as duration of a sound event/phrase).

These three music accent categories should be treated as overarching umbrella categories, to be expanded in the remainder of this chapter. "Sonic accent" refers to a sound event such as a marked pitch or local-level harmonic activity. The activity may be a specific note (pitch), a non-phrasal melodic motif/fragment, timbral quality, or simple (incomplete) harmonic motion that renders sound as a mirror of onscreen action or sound effect. This refines the formulaic combination as:

R • [Sonic or *Phrasal* or *Duration accent]-[visual accent type]*

Let "sonic" refer to any phenomenal accent according to Lerdahl and Jackendoff, where "any event at the musical surface [can give] emphasis or stress to a moment in the musical flow."[74] Because phenomenal accents can be broad, I refine possibilities into the following plugin:

[*Sonic*] = [S] = non-rhythmic emphasis of sound events

Therefore, R[S] = non-periodic sonic sync points

[74] Fred Lerdahl and Ray S. Jackendoff, *A Generative Theory of Tonal Music* (Cambridge: MIT Press, 1983), 17. Brackets, mine. Original citation: "By *phenomenal accent* we mean any event at the musical surface that gives emphasis or stress to a moment in the musical flow."

An example of rhythmic-sonic sync point *R[S]* would be Figure 2.5, where the flamingos from *Fantasia 2000* are thrust into the branch with the orchestra providing both music and sound effects for this moment. The sonic "crash" is a single timbral and pitch-related event that does not indicate a larger structure such as a phrase. It instead signifies the start of the collision while the subsequent trills are reinforced by duration and not phrase structural materials.

"Phrasal" refers to phrase-structural features such as cadences and melodic strains, as well as sentence, period, and formal musical structures. A rhythmic-phrasal (or *R[P]*) designation can highlight a structural downbeat (beginning of a measure), non-structural downbeats, or begin phrases, or an anacrusis that includes the beginnings of phrases. The most important aspect is that the R[P] function effectively synchronizes with initiation of a musical phrase. Incidental harmonic progression (I-V, V-I, or modulation) may also accompany and strengthen the musical phrase being synchronized but must not exhibit recurring behaviors or repetitions that indicate meter or hypermeter to retain the rhythmic R[x]-[y] definition; they can belong to formal delineations but must not define the form by itself. All phrasal features must follow music-theoretical functions. The sync type is described as follows:

[*Phrasal*] = [P] = phrase-structural and formal features

Therefore, R[P] = non-periodic phrasal sync points

An example of *R[P]* is Figure 2.6 featuring "Love is an Open Door" from *Frozen* (2013), where the camera shot fixates on Anna when she sings her phrase and only changes when her phrase ends and Hans enters the picture.[75] The second visual accent in the excerpt features a shot

[75] For media reference, see: DisneyMusicVEVO, "Kristen Bell, Santino Fontana - Love Is an Open Door (From "Frozen"/Sing-Along)," YouTube video, 2:05, https://youtu.be/kQDw88hEr2c?si=V-ixJYjIKXOb64Sf.

change to Hans and does not correspond to any sound. The third visual accent (character motion) is when Hans jumps onto his feet and opens his right hand. This emphatic music coincides with the beginning of Hans' phrase. Therefore, the total sequence of visual accents delineates the entire musical phrase as opposed to a singular or disparate (non-phrasal) sonic event. Even when Hans is talking (second arrow) and there is no music, the shot change to Hans anticipates the coming of his phrase and therefore supports the upcoming R[P] of the third visual accent. The image clings to the next nearest sonic event, emphasized not by the shot change, but by Hans' character (in-shot) motion. The use of in-shot motion instead of camera shots as visual accent will be expanded later in this chapter.

"Duration" refers to non-pitched musical aspects such as rhythmic patterns, velocity (accelerando/ritardando), or any temporal aspects of the music. The sync points of a certain action do not create associations with pitch, tone, timbre, or coloration. It is summarized as:

[*Duration*] = [D] = temporal synchronicities

Therefore, R[D] = non-periodic durational sync points

An example of R[D] type sync is Figure 2.7, a revisiting of the flamingo pecking scene. The pecking action, followed by incessant stomping, is synced to the music's rapid rhythmic qualities and the static harmonies do not provide a synchronized progression, nor does it create a salient melodic phrase.

Figure 2.8 illustrates how three basic film techniques inform three visual accent types. While music determines *what* is being synchronized, the visual accents determine *how* the music is synchronized and marked for attention. The type of arrows in the "arrow key" box of Figure

2.8 details what kind of visual accent is involved in the sync point. While it may seem that image would then take precedence in analysis, this is untrue since many of the case studies in this chapter adapt visual imagery to preexisting or precomposed music. Even if music was composed after the image, it is given that music is in some sense indicative of the visual direction for passages that feature mickey-mousing of any sort.

Reiterating the *R • [music accent type]-[visual accent type]* sync formula, the following section will define [visual accent type]. Indicators of [visual accents] typically occur via three general film techniques:

1. Action-stop (in-shot, usually character motion)
2. Shot-change (between-shot)
3. Camera motion (pan, zoom, etc.)

The presence of a film technique, combined with musical information, defines the type and content of visual accent in the equation. In this chapter, mickey mousing is treated more broadly as character motion, and is regarded an in-shot [visual accent type] where characters or objects on screen move in sync with one of the three musical accents. The synchronous mirroring is an action-stop visual accent marked by the start and end of motion. The primary example of this is from the "Turkey in the Straw" number from *Steamboat Willie* (1928), the first sound-synced animation where "mickey-mousing" earned its namesake.[76] Figure 2.9 excerpts the melody that is repeated six times throughout the scene, and Mickey's in-shot actions closely mirror every melodic note. Outlined in Figure 2.9, the intense focus on the in-shot motion features few to no shot changes in the first three repetitions of the song. The synchronicities here

[76] Media reference: Walt Disney Animation Studios, "Walt Disney Animation Studios' Steamboat Willie," YouTube video, 5:27, https://youtu.be/BBgghnQF6E4?si=zkXbYvViRywtL-F8.

are functional because each visually accented motion is purposed to a note or sound from the song, the purpose being to dialectically define Mickey's animated performance of "Turkey in the Straw."

The second variety of [visual accent type] is the editing rhythm, or the rate of change between shots. An accent occurs each time an image changes from one shot to another, and a functionally synchronized shot change links the rate of change to a musical function. An example of this is Figure 2.10, where each shot change emphasizes the cadential chords of "Hokori to Ogori" from *Shoujo Kageki Revue Starlight* (2018). In this sequence, *both* character motion and shot changes work together to mark synchronicities.

The third technique is camera motion. In addition to character choreographies and poses or temporal shot changes, animation often simulates the movement of camera, be it tracking shots or emphatic zooms and pans. This form of visual transition is like action-stop because there are definitive beginnings and ends of camera motion. Figure 2.11 excerpts "Honor to Us All" from *Mulan* (1998) and features clear use of camera motion synchronized to a cadence, with camera motion represented by a rightward arrow connected to the sync point indicators.[77]

In this dissertation's analytical framework, the three visual techniques interact with musical points to define the second component of rhythmic sync points. For the sake of streamlined analysis, and because the types can occur simultaneously, the accents are generalized as [*visual accents*] that define the total R[x]-[y] sync point. With these categories in mind, one accent type would naturally be more salient than the others, facilitating analysis.

[77] Media reference: Walt Disney Feature Animation, *Mulan*, directed by Barry Cook and Tony Bancroft, Walt Disney Studios, 1998, streaming video, 1:36:00, accessed June 2024, Disney Plus, https://www.disneyplus.com/movies/mulan/85wmj4hahA0B

Once a [visual accent] is identified, its function is determined by the audiovisual interaction with the music. The function forms sync types by defining the second half of the R[x]-[y] sync formula with the labels: isomorphic, iconic, and isomorphic-iconic. Shown in Figure 2.12, these categories can be combined in various ways with the three [music accent type] categories to create total R-sync types. The first type, *isomorphic* or [ISO], is a mirroring of durational structure like the R[D] music accent type. Curtis describes isomorphic audiovisuality as music and image sharing the same "shape" or "rhythm" with in-shot action and music that expresses associational fidelity (i.e., a cymbal synced to a car crash).[78] From this definition, my analytical adaptation of a purely isomorphic relation would be:

R • [Duration]-[ISO]

Where [ISO] is mirroring of durational structure.

This means that an R[D]-[ISO] sync point is a purely rhythmic unit with no regard to [sonic] or [phrasal] qualities. This relation focuses both music and image on the rhythmic, non-pitched qualities of the music and action. An example of this is Figure 2.13 from *Bugs and Daffy's Carnival of the Animals*, another animated interpretation of Camille Saint-Saëns' "Finale."[79] The musical passage features the main theme, but direct shot changes to imagery of violins and horns occur on each beat with little expression or painting of the melodic contour evident in the *Fantasia 2000* adaptation. While the excerpt is a musical phrase, the sync points are purely rhythmic because they only mirror the temporal aspects of the phrase. In the same vein, it is generally maintained that [ISO] visual accentuation rarely synchronizes with the R[P] relation.

[78] Curtis, 201.
[79] This analysis also references another DVD video. See: Warner Bros Pictures, "Bugs and Daffy's Carnival of the Animals," in *Looney Tunes Golden Collection: Volume Five*, directed by Chuck Jones (Burbank: Warner Home Video, 2007), DVD, 417 mins.

It is also possible to have an *R • [Sonic]-[ISO]* relation labelled R[S]-[ISO]. In this relation, temporal and durational aspects of the music take precedence over all other musical features (i.e., pitch, timbre, etc.). Figure 2.14 is an excerpt from *Bugs and Daffy* where the piano part bears timbral association with the visually accented lines and keys in the animation. Instead of emphasizing the notes being played, colors appear randomly but in rhythm with the music. Therefore, this sync type focuses on sound source and timbre (the "what") by rhythmically punctuating the sketch and coloring of a keyboard (the "how").

The second visual accent type is *iconic*, or [ICO], the literal or abstract pictorial association to non-durational musical aspects such as timbre, pitch, melodies, fragments, motives, and so forth. What sets an R[S]-[ICO] relation apart from an R[S]-[ISO] is that both sound and image work together in a non-rhythmic way to create a composite meaning that propels the narrative. For example, Figure 2.15's sequence uses in-shot, action-stop accents to mark the scalar runs in the passage. The rightward, dotted arrow line represents the emphasis and presence of in-shot choreographies. The flamingo's choreography and animated intention is generally understood to be "dance," where he moves in rhythm that is not closely mirrored to *every* note in the melody. Instead, the action-stop of the yoyo and flamingo—every motion and pose—sync with the piano's scalar runs. Although the yoyo flamingo's action-stop motions are rhythmic in a sense that a phenomenal accent syncs at the downbeat of m. 27 and is paralleled in the following m. 28, the sync points are [ICO] (and *not* [ISO]) because the flamingo's complete actions "stand in" for the scale. This propels the narrative and adds an extra layer of meaning instead of purely mirroring durational values.

An R[P]-[ICO] connects a phrase or formal unit to express synchronized narratives. What separates this from an R[S]-[ICO] relation is that a recognizable melodic phrase (instead of a

scalar or virtuosic passage) serves as the musical content emphasized by synchronous [ICO] accents. Figure 2.16 illustrates an R[P]-[ICO] from the *Bugs and Daffy* adaptation of the "Finale." The various orchestra instruments are visually represented in the live action shot of the orchestra Bugs and Daffy perform with. However, the sound is iconic because it refers to "orchestra" without having to show the in-shot performance of each instrument explicitly and accurately. All orchestra parts are inferred by the image content, even if the pictured instruments are not currently sounding in the music. The abstracted idea of "orchestra sounds" is represented through visually accented images (iconic), and the attribution of said sound to the final cadence, the [P] phrase, is what relegates the sync point (a shot change) to the R[P]-[ICO] label.

Bugs and Daffy also features R[D]-[ICO] in Figure 2.17. The ideas of "jumping" and "kangaroos" are visually represented in animation that is synchronized to the grace-note piano patterns in the music. While the jumping mirrors durations of the music, the action-stop of the large kangaroo still lacks the hyper-precision of an R[D]-[ISO] relation.

It is possible for both [ISO] and [ICO] to occur at the same time, resulting in the following formulaic combination:

R(x)-ISO • R(x)-ICO = R[x]-[I^2], where (x) stands for [S], [P], or [D]

The label R[x]-[I^2] identifies a musical feature that is emphasized by both rhythmic (isomorphic) and non-rhythmic (iconic) visual accents. Figure 2.18 presents various forms of R[x]-[I^2] sync.

R[S]-[I^2] in Figure 2.18a excerpts *Flamingos* where the common flamingos slam into a tree branch and suffer a spiraling recoil in cartoon fashion. The sonic event, an orchestral clap and tremolo underlying the main melodic action, serves as a written-in sound effect for the action. The durational aspect of the action (isomorphic) pairs with the timbral quality of the

sound effect, creating an action-based narrative (iconic) for the birds. These [visual accents] features punctuate the [S] quality of the music, or non-rhythmic aspects of this sound event. It is not a purely rhythmic event because the effect relies on timbral qualities and difference to sync to a different layer of the music score.

The R[P]-[I^2] excerpt in Figure 2.18b examines the "march," "jump," "descend" sequence from *Flamingos*, where the entire phrase is mirrored rhythmically (isomorphic) and narratively (iconic). The pitch contour of the melody is clearly visualized by the flamingos' action-stop motion, and the rhythmic marching is precisely mickey-moused to the eighth note rhythm of the repeated E's. Even the mid-air suspension between jump and landing lasts the duration of the quarter note trills. The abundance of up-down and dotted rightward arrows illustrates the level in-shot synchronicity the visual accents bear with the musical materials. The vertical arrows mark every sync point in the excerpt, while the rightward dotted arrow represents constant character motion unfolding in real time.

The R[D]-[I^2] example of Figure 2.18c reexamines the "sun ray" motif from the opening of *Flamingos*. The visual accents have a clear action-stop with the sun rays appearing and holding their completed "poses." The duration is mirrored by the glowing rays, and the narrative of "sun rays" is expressed by the in-shot motion. Each idea lasts the duration of a musical feature such as the piano tremolo ("sun rays"), creating new composite associations. As [D] stands for non-pitched synchronicities, only the duration is the primary aspect of the synced music here. The isomorphic aspect is the mirroring of duration, while the iconic aspect connects specific animated actions to respective musical features.

These combinations for the *R • [Sonic* or *Phrasal* or *Duration accent]-[visual accent type]* formula identify the building blocks of synchronized, animated music sequences. Much

like Roman Numeral analysis, the labels are not meant to pinhole but to bring forth and acknowledge the various audiovisual techniques that work together to make a scene happen.

2.3: Analytical Applications

Applying sync point analysis gives insight into how an animated scene creates narrative on the musical and visual fronts. Two different streams of information come together at synchronized moments to mark certain information that adds up to a composite audiovisual presentation. Labelling R-sync types not only provides a play-by-play on how synchronicity or mickey-mousing is technically achieved, but also determines which ones are functional to audiovisual storytelling. Just like pure musical analysis, it may reveal overarching themes or stylistic choices for an animated sequence.

The focal case study, *Flamingos* from *Fantasia 2000*, is revisited to demonstrate this purpose. This fully animated sequence connects an audiovisual narrative to the "Finale" of Camille Saint-Saëns' *Carnival of the Animals*. The visual story and musical surface work together to follow a humorous contest between a free-spirited yoyo flamingo and his conformist flamingo friends. Although the yoyo flamingo meets opposition from his friends who engage in cartoon violence such as pecking, stomping, and kicking, his desire to simply have fun leads him to victory at the end of the narrative.

In line with Disney animations' penchant for savvy sound synchronization techniques, *Flamingos* mostly features instances of R[x]-[I^2] varieties. This calls attention to the complexities of sync point types in this study. Because *Flamingos* adapts an orchestral work, there is no "singing voice," thus positioning analysis towards abstract contexts and a reliance on sync points to determine function.

The variety of R[D] sync point types are loosely attributable to non-characterizing onscreen action such as camera motion. From the onset of Figure 2.19a, “sun rays” is the first sync point labelled R[D]-[I^2]. Arrows in the score represent the visual accent rhythm and each sync point that attaches itself to a musical note or feature. Here, the R[D]-[I^2] plays a role in setting the stage via sun rays visually accented by in-shot motion. An R[D]-[I^2] sync point shown in Figure 2.19b also appears as a “jump” motion within the main theme that primarily reflects the duration of the note.

Additionally, instances of R[D]-[ISO] in 2.19c (m. 26) and 2.19d (mm. 81–83) synchronize with the camera motion by reflecting the duration of the measure without mirroring the pitch or pitch contour features found in the music. The first instance of R[D]-[ISO] in m. 26 (Fig. 2.19c) brings visual focus to the counternarrative and nonconformist character of the story: a late-rising, yoyo flamingo. Synced to the fourth beat with an emphatic chordal strike, the camera motion blurs across the screen. It is purely timed, and the blurring does not provide a text-painting, pictorial association of non-rhythmic contours and is therefore, [ISO]. In the same manner, camera motion brings in an R[D]-[ISO] before the closing of the “Finale.” Figure 2.19e highlights how near the end of the movement, an R[D]-[ICO] sync usurps the bassline, after the common pink flamingos “defeat” the yoyo flamingo. To their dismay, a yoyo pops out from the water and wraps around them within the duration of the last three ascending bass notes.

By means of in-shot motion, R[S] sync types, particularly R[S]-[I^2] points, present most of the audiovisual action and characterization of the flamingos. The entrance of the “waking up” adds another sync point, an R[S]-[I^2] in Figure 2.20a, that signifies the [Sonic] quality of the glissandos with swirling birds [ICO] and camera panning [ISO] that is timed to each ascending piano pattern. In Figure 2.20b, the beginning of every main theme statement as that in m. 11, is

synced to R[S]-[I^2] points, depicting the organization and conformity of the common pink flamingos. In Figure 2.20c, R[S]-[I^2] takes the form of a sound effect in m. 58.

When the yoyo flamingo begins to dance in the excerpted passage of Figure 2.20d, the sync points align with the [durational] scalar qualities of each four-note scalar pattern but serves more narrative purposes than a purely durational sync point. This leads to the R[S]-[ICO] interpretation for the sync points in mm. 27–34 with some interspersed [ICO] moments such as. Here, the R[S] type sync points emphasize the yoyo flamingo's playfulness through mirroring yoyo action to the scalar passages. In Figure 2.20e, R[S] sync mirrors the scalar passage where the pink flamingos rise from the water to move side to side and kick the yoyo flamingo into the air. All these in-shot actions not only invigorate the animated sequence but also visually emphasize the common vs. yoyo flamingo narrative through musically mirrored action.

Though instances of R[P]-[I^2] are fewer than the R[D] and R[S] varieties, R[P] sync points highlight some small ascending phrases shown in Figure 2.21a and two major musical features of the movement: the main theme (Fig. 2.21b) and the final cadence (Fig. 2.21c).[80]

The first theme statement shown in Figure 2.21b was categorized previously as an R[S]-[I^2] leading into an R[D]-[I^2] "jump" and R[S]-[ICO] "descent." This can alternatively be described and reinterpreted as an R[P]-[I^2] complex that reflects a musical theme ("phrasal" = [P]) while simultaneously subsuming the three sync point Units A-C annotated in the example. Figure 2.21b demonstrates how an R[P]-[I^2] complex can encompass different sync point types under the umbrella of a phrasal structure. In the thematic statement, muddy mickey-mousing reflects the rhythm, articulation, and pitch contour of the main melody, albeit slightly mismatched in timing the "jump" and "descent." Unit A presents four R[S]-[I^2] sync points where

[80] Sync points and cadential passages will be discussed in depth in the next chapter.

the flamingos march in a static motion that mirrors the pitch content, while rhythmically moving to the eighth note rhythms. There is slight modification to the R[D]-[I^2] in Unit B because the "jump" matches the duration but not raised note motion of the pitch movement. The Unit C that encompasses sync points 6–11 returns to an R[S] type but is R[S]-[ICO] because the descending motion is more iconic than the precise mickey mousing offered by an [I^2] type. As a matter of fact, the ascending D#-E motion in the high woodwinds is matched to the visual descent at first. Only at sync point 8 does the descent match the pitch contour.

Categorizing and comparing sync points also provides data for narrative and hermeneutic interpretations. There is a predominantly [I^2] association with the common flamingos presented in the opening, the melody, and the flamingos' return. The association differentiates in retrospect any non-[I^2] that have appeared, and correlates said non-[I^2] to the yoyo flamingo. The result is a narrative split between the common flamingos ([I^2]) and yoyo flamingo (non-[I^2]) based on timbre (piano vs. orchestra) and sync point types. The split associations are reinforced in areas like Figure 2.22a (m. 39), when the yoyo flamingo maintains sync with the piano line while the common flamingos' sync with the orchestral timbre. The composite R[S]-[I^2], synced to the rising bassline, reinforces [I^2] whenever the pink flamingos are present. The passage closes out with an increase of sync point types that transition back to a restatement of the main melody, set with the same R[P]-[I^2] sync point type as the prior main theme presentation (mm. 11–26). Again, in Figure 2.22b (m. 62), the common flamingos reestablish their presence by synchronizing to the rising strings, while the yoyo flamingo hijacks the sync point focus shortly after (m. 66) with non [I^2] variations. Towards the end, the common flamingos eventually reassert themselves with R[S]-[I^2] sync to the orchestra (m. 76) and maintain this when they stomp the yoyo flamingo into the water (m. 78). Ironically, the yoyo flamingo bursts from the

water at the end and claims victory with a cadential passage synced with R[P]- [I^2] type. This sync point type "switch" both thwarts conditioned expectations set up throughout the audiovisual narrative and captures the ironic nature of the story adaptation.[81]

The highlight of *Flamingos* is that the animation maximizes visual accentuation through use of R[x]- [I^2] and primarily in-shot and character motion. This is a deliberate choice in the direction of this adaptation. However, as the following comparative analysis demonstrates, between-shot motion is another valid form of audiovisual storytelling.

2.4: Comparative Analysis

In *Flamingos*, rhythmic sync points advance or paint an overall narrative. The sync point's temporal placements, the emphasized visual content, and highlighted musical features all work together to create and allow hermeneutic readings of an emergent narrative meaning. Aside from narrative interpretations, sync point analyses allow comparisons of styles among animation studios. As an example, the following case study compares the *Flamingos* (2000) adaptation of *The Carnival of the Animals* "Finale" with that of *Bugs and Daffy's Carnival of the Animals* (1976) by Warner Bros. Analyses reveal that the former (Disney) focuses on fluid synchronicities between image, in-shot motion, and music while the latter (Warner Bros.) focuses on synchronizing music to slapstick humor, shot changes, and more static animations.

Despite the irony, *Flamingos* presents a nuanced and serious storytelling bent that eliminates any hints of an artist or fictionalization.[82] This invites suspension of disbelief and immerses viewers into the narrative. *Bugs and Daffy's* narrative takes the opposite approach,

[81] This yoyo sequence will be expanded on in the next chapter.
[82] The absence of the artist's hand or any indication of live action is believed to add to suspension of disbelief when viewing animation.

combining animation images with live action shots of the orchestra and conductor Michael Tilson Thomas. There is little to no audiovisual storytelling like that of *Flamingos*. Instead, this adaptation juxtaposes the performance of a live orchestra with abstract and iconic animations of orchestral instruments and animals, save for one instance of Bugs and Daffy playing the piano.

The additional dialectic between live and animation styles in the adaptation also brings forth animation's ability to use editing techniques to communicate more abstract isomorphic and iconic visual information. The tone for *Bugs and Daffy's* is humorous and slapstick, with less attention to narrative and more to abstract animations. The presence of a live orchestra lessens the narrative weight for *Bugs and Daffy's*, making room for more sync point possibilities without overtly relying on [I^2] types to evoke hyperrealism.

Bugs and Daffy's is highly saturated with R[D]-[ICO] sync point types. This insinuates that there is less focus on ascribing meaning to certain notes and musical features, resulting in a more abstract [ICO] experience. The opening is stark contrast to *Flamingos*. Salient with R[D]-[ICO] synchronicity, shot changes are the predominant visual accents. The visual narrative of the live action is not as manufactured as animation, and the general idea is understood as "conductor" directing and "orchestra" performing the music.

Whereas the visual aspects of R[x]-[I^2] were the constant for *Flamingos*, the musical-durational aspects of R[D]-[y] are central to *Bugs and Daffy's*. The Warner Bros. adaptation utilizes R[D] as the salient audiovisual relation, demonstrating how an animated sequence can establish a good portion of its audiovisual narrative through focus on a minimal set of sync point types. Excerpted in Figure 2.23a, the lack of animation reduces the amount of synchronous motion between orchestra musicians and the musical material, resulting in an R[D]-[ICO] sync for the first twenty-six measures of the "Finale," whereby the timbral composite of "orchestra" is

represented. In Figure 2.23a, the opening nine bars are shown. The shot changes are not mickey-moused and instead focus on reinforcing the conductor and orchestra narrative. Unlike *Flamingos*, both opening tremolos and following three-note motif are not assigned a narrative function in the *Bugs and Daffy's* adaptation. While the opening may seem to accent an R[S] type synchronicity, this analysis reads it as R[D] because the shot changes do not emphasize the timbre of the woodwinds or strings, even changing shots before the strings pick up the three-note motif. Without special emphasis, there is no direct assignment of an in-shot content to sound, therefore making the changes more durational than sonic. R[D]-[ICO] is also prominent when the camera cuts back to the live action shots of the conductor and orchestra towards the end of the "performance."

R[D] sync also depicts in-shot action such as that in 2.23b. Here, an instance of R[D]-[ISO] appears in m. 35, at the initial formation of the "Finale" lettering in an animated title card. In humorous fashion, the title of the song does not appear until this point, having passed the opening and first statement of the theme. In this sequence, the different colors in the title card lettering appear non-periodically and do not immediately attribute themselves to any timbre, instrument, or sonic feature. Therefore, the sequence is strictly duration-based, earning the R[D]-[ISO] label. The return of the main theme features R[D]-[ISO] visualized through abstract shapes and cuts of a talking mouth, a lion, fish, and fossils are more durationally prominent than they are iconic of associating meaning to the sound they coincide with. Example 2.23c excerpts R[D]-[ISO] moments found in the farm chicken sequence where music and image loosely represent "pecking" without timbral associations. Likewise, the horn imagery in Figure 2.23d features no sonic fidelity to the ongoing strings and reiterates the R[D]-[ISO] sync points.

Ironically, Bugs and Daffy appear only once in the entire sequence and do not feature R[D] type sync in their short-lived appearance. Instead, animation comes to the center in Figure 2.24 (m. 27), where the live action focus cuts to Bugs Bunny performing the piano part with R[S]-[ICO] sync. This is like the R[S]-[ICO] in *Flamingos* because Bugs' piano playing (and later, Daffy's in m. 31), expresses the qualities of piano-playing but does not precisely mirror the note-for-note piano playing. The piano is implied by the back view and black outline, but the keys and finger motion are not clearly animated to qualify for the mickey-mousing provided by an R[D] sync. The sync points therefore align with the durational qualities of the scalar passage for all three cuts to Bugs, the conductor, and Daffy. These are [ICO] in a different manner than in *Flamingos*. Instead of representing a yo-yo flamingo "dance," the [ICO] is less literal, standing in for timbre or the association of "piano" and "orchestra" to abstracted or inferred images that lack precise mickey-mousing. Therefore, the rate of shot changes is slower than *Flamingos*, allowing slightly more time for abstract associations.

There is only one instance of R[P]-[ICO] to accent and focus on the climactic passage leading into the final cadence, a common convention that will be discussed in a later chapter. In *Bugs and Daffy's*, the main theme does not receive an R[P] treatment as it did in *Flamingos*. Instead of a continuous narrative, the presence of shot changes (or lack thereof in live action portions) does not present mickey mousing in any capacity. While the theme presentation may feature rhythmic change rates like that of *Flamingos* where the musical phrase is synced at many locations within the bar, the disjoint visual content and unregulated sync rates differ from *Flamingos*' consistent mirroring of musical articulations and contours.

Aside from labelling frequent sync types, sync point analysis also pinpoints areas of interest where two different types of sync may coincide. In Figure 2.25, two kangaroos enter and

present a complex sequence as there are *two* types of sync points. The main R[D]-[ICO] is still maintained by the general imagery and the larger hopping kangaroo represented by the rhythms in the top line. However, the smaller, quicker moving kangaroo creates a simultaneous R[S]-[I^2] sync with the grace note piano passage. This is the only instance of R[S]-[I^2], albeit not purely in focus as would be the sync points in *Flamingos*.

Exemplified in Figure 2.26, an in-depth and analytical comparison of two different adaptations of the same pre-composed musical work gives technical insight into the stylistic differences. For *Bugs and Daffy's*, the narrative does not go any deeper beyond visually emphasizing the orchestra, conductor, and abstract animation triangle. The abstract animations feel kaleidoscopic, offering tidbits of vibrant imagery and non-narrative visual attractions facilitated by the salience of R[D] type syncs. The switches between R[D]-[ISO] and R[D]-[ICO] throughout the adaptation balance duration-based and meaning-based visual accents, enough to avoid the hyper-detailed narrative of the *Flamingos* adaptation.

The supposed main characters of the adaptation, Bugs and Daffy, are shown to be part of the orchestra, receiving cues from the conductor, and they ironically do not participate in any of the other animation sequences. Any appearances of other sync types, including the R[S]-[I^2] at the kangaroo sequence, are kept to a minimum and are often bookended by the predominant R[D] sync types. Using mostly durational properties R[D] and [ISO], *Bugs and Daffy's* continues to adhere to the slapstick Warner Bros. animation model. The inclusion of the live conductor and orchestra is also ironic, being vaguely iconic of the sonic fidelity while being absent of animated mickey-mousing. Released in 1976, *Bugs and Daffy's* interactive live-action and animation bring forth the Warner Bros. style without committing to faithfully animating a perfect piano performance on Bugs' and Daffy's parts. The live action portions contribute to the performance

bent of the adaptation. Abstract visuals help music and image work together without any of the components being subservient to the other.

Flamingos takes an opposite approach, eliminating the artists' presence (both musical and artistic) in favor of a storytelling narrative about the yo-yo and common flamingos. R[x]-[I^2] sync types like that in Figure 2.26a strengthen the commitment to creating a serious narrative and audiences' suspension of disbelief. Whereas the R[D]-[x] musical accents anchor *Bugs and Daffy's* antics excerpted in Figures 2.26a and 2.26b, the [I^2] visual accents of the R[x]-[I^2] sync type visually anchor the musical nuances highlighted by Disney's *Flamingos* adaptations of the same passages. Through this, features such as musical phrases and instrumental timbre are more emphatic in the Disney adaptation. It is almost as if the music determines the actions of the flamingos. However, both music and image are of equal grounds because, while the music influences the mickey-mousing flamingos, the visual accents emphasize the motion, characterization, and narrative drama that unfolds, leading to new meanings. After all, flamingos are not featured in Saint-Saëns' original movements and are Disney's original addition to the animal carnival.

2.5: Conclusions

Rhythmic sync points defined by the formula *R • [music accent type]-[visual accent type] = R[x]-[y]* allow analysis of local-level events that demonstrate the what and how of audiovisual relationships. It brings forth the components involved in moving a narrative from one point to another. Identifying sonic, phrasal, or durational events in music shows *what* parts of music are being synchronized with image. Likewise, identifying the visual accents emphasized by character, camera, or shot-change effects shows *how* the music is emphasized. Together,

synchronized moments between audio and visual create associations and meanings that may even highlight moments that are traditionally non-functional in music or film theories. By comparing *Flamingos* as well as *Bugs and Daffy's*, the two main case studies of this chapter, it is viable that sync point analysis illustrates not only the inner workings of each adaptation but also highlights the studios' different animation styles.

I tapped into Scott Curtis's principles and critique of traditional film music analysis to outline my methodology. For analyses to happen, the soundtrack was generalized, meaning that sound effects written in music and the synchronized ("mickey-moused") musical material were treated as a single entity. My focus on film rhythm ignores non-musical audio such as dialogue, voice-over narration, and sound effects. This allows analysis of moments where sonic events in the music were visually adapted into sound effects (as in *Flamingos*). The diegetic and non-diegetic distinctions are de-emphasized in identifying functions, prioritizing what is musically presented instead of what is being heard by whom. This generalization of musical source and reception is made in favor of studying the immediate effects of synchronous audiovisual rhythm. The cyclical relationship between audio and image, one informing the other, alleviates the image/sound hierarchy, opening room for interpretations of emergent meanings through sync point interactions.

As Chapter 3 will demonstrate, there are two possible subsets of rhythmic sync types: (1) the *audiovisual cadence* as a subset of R[P]-[y] sync and (2) the *vococentric* editing rhythm as a subset of iconic, R[x]-[ICO], sync types. These moments magnify sync type functions and illustrate how specific audiovisual relations can be.

Chapter 4 will demonstrate how although rhythmic sync points occur at the local-most level, their functions can be projected onto larger, metric levels—or sync type supersets.

However, the most basic sync points must be identified to make larger claims on metric and hypermetric levels. Elevated to larger levels, the local R-sync types define and energize the processes that move audiovisual narratives from one point to another, from Point A to Point B.

Chapter 3

Rhythmic Sync Subsets: Audiovisual Cadences and Vococentricism

In their comprehensive writings on animation, two Disney animators quote Walt Disney's philosophy that music is "a study of rhythm, the dance" that pervades everyday lives.[83] This quote underscores how Disney's animated features are notably synchronized with the film score at marked moments. This chapter introduces two specific audiovisual markers: the *audiovisual cadence* and the *vococentric shot change*. These two markers strengthen sync point reading and are "subsets" of the rhythmic sync equation R[x]-[y] from the previous chapter. Defining these two subsets seeks to further illustrate how audiovisual synchronicity expresses, cinesthesic synesthesia, or how animation represents sound.[84]

This chapter also demonstrates how my analytical framework expands Soviet film director Sergei Eisenstein's audiovisual analytical graph from *The Film Sense* (1943). As I will illustrate, evolving animation techniques move beyond Eisenstein's original graph, prompting me to add substitutions of traditional editing cuts with rhythmic sync analyses developed in the previous chapter and nuanced with their subsets defined in this present chapter.

My analyses incorporate ideas from Eisenstein, animation principles laid out by animators Frank Thomas and Ollie Johnston, and music theories of rhythm and meter by William Rothstein,

---

[83] Frank Thomas and Ollie Johnston, *The Illusion of Life: Disney Animation* (New York: Hyperion, 1995), 285.

[84] The term cinesthesia is adapted from Vivian Sobchak's term "cinesthetic" to represent how animation visually feeds viewers a form of produced synesthesia. Used in Gregory Flaxman, "Once More, with Feeling: Cinema and Cinesthesia," in *SubStance* 45, no. 3 (2016): 174-189, muse.jhu.edu/article/634224.

Flaxman credits Vivian Sobchack for the term cinesthetic, which explores the bodily senses engaged or not engaged in a film viewing experience. For more, see: Vivian Sobchak, "What My Fingers Knew: The Cinesthetic Subject, or Vision in the Flesh," in *Carnal Thoughts: Embodiment and Moving Image Culture* (Berkeley: University of California Press, 2004), 53–84.

Sobchak's ideas are also further expounded in a thesis by Joachim Strand, who uses Sobchak's cinesthetics as a framework for determining the felt and "physiological" sensations of experience musical multimedia. For more, see: Joachim Strand, "The Cinesthetic Montage of Music-Video: Hearing the Image and Seeing the Sound," MCA Thesis, Curtin University of Technology, 2006.

Justin London, and Fred Lerdahl & Ray Jackendoff. Following Thomas and Johnston's twelve principles of animation and Wilfred Jackson's bar sheet ratios of animated frames to musical beats, I present an interdisciplinary fusion among music, film, and animation theories to explore a cadence type that follows music-theoretical conventions and is visually reinforced. In doing so, this chapter pushes beyond the issues of terminology presented in Chapter 2, to delve into deeper aspects of function such as cadential motion and vococentric phrasing.

3.1: The Audiovisual Cadence

This chapter introduces the concept of *audiovisual cadence, which* visually reinforces the end of musical phrases, sections, or formal structures found in film musical sequences. From a music-theoretical standpoint, cadences effect "closure" on musical processes that range from phrasal to formal structures, as defined by William Caplin.[85] According to William Rothstein, this closure is typically preempted by "drive" or "marked acceleration" towards the ends of phrases in terms of harmonic rhythm and melodic action.[86] The abundance of cadential motion in animated musicals warrants application of my analytical framework to audiovisual analysis that investigates recurring synchronicities between editing rhythm and musical cadences. As referenced in the literature review and Chapter 2, Frank Lehman attributes "phrasal-" or "cadential mickey-mousing" to audiovisual events that repeatedly synchronize with cadential

[85] William Caplin, *Analyzing Classical Form: An approach for the classroom* (New York: Oxford University Press, 2013), 55. I also adapt Caplin's differentiation between cadences, harmonic and melodic closure of a phrase, and codetta, closing material that *follows* the cadence to reinforce the tonic key. As noted by Caplin on p. 153, codetta may be "supported harmonically by a cadential progression" and may have "melodic components" that resemble the original PAC, but it should be distinguished based on music-theoretical functions. In other words, cadences, if evident in codetta, should be restatements of an earlier authentic cadence.

[86] William Rothstein, *Phrase Rhythm in Tonal Music* (New York: Schirmer Books, 1989), 22–25. Rothstein notes that there is typically a "drive" of "marked acceleration" towards the ends of phrases. He writes of a "pre-cadential acceleration" that "signal the coming cadence, often in conjunction with other changes…" If I elevate this definition to the audiovisual level, "other changes" can include the acceleration of shot changes, or filmic rhythm.

actions and harmonies.[87] Expanding the idea of "cadential mickey-mousing," I seek to define *what* (audio) and *how* (visual) moments of audiovisual closure are created through sync points.

In the previous chapter, R[P]-[y] sync points are defined as non-periodic visual accents that coincide with phrase-structural and formal features found in the music.[88] The audiovisual cadence, hereafter [AVC], is a subset of the R[P]-[y] sync types because it marks a cadential formal feature within a combination of units that includes the cadential point and a typical climactic passage that leads into a cadence. The simple plugin definition for an audiovisual cadence is as follows:

[AVC] $\subseteq$ **R[P]-[y]** sync point types

Where [y] = [ICO] or [I^2], and [y] $\neq$ [ISO][89]

This formula defines the audiovisual cadence [AVC] as a subset of sync points that rhythmically emphasize the phrase-formal musical cadence. The [y], or visual accent, aspect of the formula tells that there must be *iconic* [ICO] or [I^2] visual reinforcement. However, an [AVC] is not individually [ISO]-type because the strictly durational properties of an [ISO] visual accent potentially ignore the nature of phrase-structure and cuts through traditional formal structures.

---

[87] Frank Lehman, "Hollywood Cadences: Music and the Structure of Cinematic Expectation," in *Music Theory Online* 19, no. 4 (2013), doi: 10.30535/mto.19.4.2. Lehman's focus is on the harmonic aspects of cadences in film music and how they are visually supported by onscreen visuals, whether in editing, shot content, or composition. He also attributes phrasal mickey-mousing to film editing, where he writes: "With phrasal mickey-mousing, musical units are made to correspond to film editing rhythms through manipulation of phrase tempi and durations (and, rarely, after-the-fact edits to match precomposed score)."

[88] In the R[P]-[y] relation, the R stands for "Rhythmic," the [P] for "Phrasal" structures, and [y] is a placeholder for any relevant visual accent types.

[89] As explained in the previous chapter, there is no R[P]-[ISO] because [ISO] highlights rhythmic durations over pitch-based functions.

As part of a cadential structure, the [AVC] rarely acts as an individual event. It is usually reinforced by a phrasal passage that accelerates the rate of visual and musical action, and can be defined as:

[AVC] complex = a combination of units R[x]-[y] + ([AVC] $\subseteq$ R[P]-[y])

Here, the [AVC] complex is a result of an R[x]-[y] phrasal unit that *includes* the [AVC] instance that is a subset of the R[P]-[y] sync point relation. While the R[x] component could mean that any sonic, phrasal, or durational aspect of the music leads into an [AVC], it would most often be an R[P]-[y] type that subsumes the [AVC] cadence within its phrase. Therefore the ([AVC] $\subseteq$ R[P]-[y]) clause of the formula defines how the [AVC] will always be a subset of the R[P]-[y] sync point relation. Just as in Chapter 2, the [y] variable is a placeholder for any of the [ICO] or [I^2] forms of visual accentuation.

The [AVC] instance itself follows traditional music theoretical cadence types and can be specified as half or authentic (perfect and imperfect). To clarify the type of cadence being synced, the [AVC] is refined as follows:

- Perfect authentic cadence: [P-AVC]
- Imperfect authentic cadence: [I-AVC]
- Half cadence: [H-AVC]
- Deceptive cadence: [D-AVC]

As the following figures illustrate, an [AVC] complex will include the main [AVC] which is a subset of an R[P]-[y] passage as well as R[P] phrasal material that leads up to the cadence.

Figure 3.1 presents three instances of audiovisual cadences where shot changes coincide with final cadences, and the remaining case studies illustrate variations of the cadence at sectional endings and with differing visual techniques. These figures excerpt concise versions of [AVC] complexes from three different films. All three examples mark the moment where the cadence occurs with the label "[AVC]." This precise moment is part of an R[P]-[AVC] complex that includes an R[P]-[I^2] passage *and* the precise [AVC] moment. In Figure 3.1a, the light gray annotations indicate what each rhythmic sync point would be *individually*. However, the existence of the [AVC] at the end of the complex *reinforms* the functions of the individual R[S]-[I^2] and *transforms* the functions of the initiating sync point from singular [Sonic] events to parts of a whole [Phrasal] passage. To clarify, the *initiating* R[P] will always be the first sync point of any given R[P]-[AVC] passage. While individual sync points can be labeled after the initiating R[P]-[I^2], they are greyed out or omitted in the annotations for clarity.

As presented in Chapter 2, shot changes that reflect Eisenstein's original montage theories remain primary indicators of visual accent. Figure 3.1a excerpts the opening number, "Where You Are" from the titular character's movie, *Moana* (2016).[90] Although Moana's song features constant camera and character motion, the ending cadential passage's increased cutting between shots emphasizes the final cadence progression ii-III-IV-V-I in E Major. It also correlates with the melodic repetition of each measure that emphasizes and reemphasizes the titular lyrics "where you are." The acceleration of shot changes coincides with the quickening harmonic rhythm, matching the downbeat of each musical measure.

Annotated as an R[P]-[I^2] unit, this [AVC] complex includes what would individually be R[S] sync point types. But because of the [AVC] in the third measure and repeating motif, the

[90] Media reference: DisneyMusicVEVO, "Where You Are (From "Moana"/Sing-Along)." YouTube video, 3:16, https://youtu.be/RTWhvp_OD6s?si=bhUeaIlXKei49Ll0.

complex is recognized as a cadential phrase that ends the song, therefore transforming the R[S] status to the R[P]-[I^2] because it fulfills synchronicity to (1) a non-periodic phrasal event and (2) a formal cadential feature, and (3) it tracks a functional cadential harmonic progression as opposed to a small harmonic motion.

This [AVC] complex is fascinating because the musical rhythm repeats and accelerates to the end, creating an expectation for a musical cadence that needs only confirmation. This confirmation is fulfilled at the [AVC] point in Figure 3.1a. This is slightly different on the audiovisual level, where both music and image must function in tandem and in real time. According to Chapter 2's framework, the R[S] sync label would remain an R[S] sync had there been no retrospective reinforming of an R[P] passage. The simultaneous expectation of cadencing in the musical part, marked by the downbeat-oriented visual accents functionally support the [AVC] cadence on all three musical, visual, and audiovisual levels.

Figures 3.1b and 3.1c excerpt "Hokori to Ogori" (trans. "Pride and Arrogance"), a focal musical number in the Japanese anime series *Revue Starlight* (2018).[91] There are two moments of synchronized cadences in this sequence. The first (Fig. 3.1b) appears at the end of the introductory section of the musical scene, synchronizing cuts to the iv9–iiø7–V–I progression in E minor. The second example (Fig. 3.1c) is the final cadence featuring an iiø7–V9–I progression in G minor. Figure 3.1b closes off a preceding section, cadencing into a character solo, while Figure 3.1c ends the entire musical sequence, mirroring the dynamic visuals. Like Figure 3.1a,

[91] Due to licensing, *Revue Starlight* (in Japanese: 少女☆歌劇 レヴュースタァライト) has been licensed for stream on different streaming websites (such as HiDIVE and the now inactive VRV), at different times. As of Junes 2024, the series is available on HiDIVE: https://www.hidive.com/season/20527. The most stable access is currently through DVD or Blu-Ray releases. The song "Hokori to Ogori" is inserted into Episode 3 of the series.

Media reference: Kinema Citrus, *少女☆歌劇 レヴュースタァライト* (Shōjo Kageki Revyū Sutāraito), directed by Tomohiro Furukawa, Sentai Filmworks, 2018, streaming video, 25:26, accessed June 2024, HiDIVE, https://www.hidive.com/season/20527.

both excerpts feature an [AVC] that reinforms erstwhile [Sonic] events and transforms the initiating [Phrasal] sync point from an [S] to [P] function. Figure 3.1c differs from 3.1a and 3.1b in featuring an additional sync point after the [AVC]. This point is, however, included in the R[P]-[AVC] complex as octave reinforcement of both the [AVC] in m. 2 (Fig. 3.1c) and the tonic key of G minor.

Sometimes, an R[P]-[y] + [AVC] passage is not as melodic as the *Moana* example. It may also include brief codetta-like materials such as in Figure 3.2. Both the Disney *Flamingos* and WB *Bugs and Daffy's* versions demonstrate different approaches to synchronizing the chordal materials that reinforce the final cadence at the downbeat of m. 3. This passage is approached by a chromatic ascent to the octave C on the downbeat of m. 1, averting a strong [P-AVC]. This fleeting moment is accented by both the Disney and WB versions.

Disney's higher frequencies of sync points reinforce a reading of an [AVC] at the downbeat of m. 3. In the Disney version, the Yoyo Flamingo asserts victory by bursting from the water at the first sync point, an R[P]-[I^2], and flinging yoyos across the screen. This is the initiating sync point for the onscreen, in-shot actions. Should this be the case, the chordal toggle between I-V-I (C and G major harmonies) pushes towards a *correction* of the weak [AVC] to the more solid root-position V-I [P-AVC] in m. 3 of the excerpt.

According to the sync profile of the WB *Bugs and Daffy's* version, the downbeat of m. 1 could be a weak [AVC] cadence where the global tonic key of C Major is supported by a chordal codetta of sorts. The lack of sync points in mm. 3–4 weakens the case for a [P-AVC] at the audiovisual level because there is no visual accentuation synced to the musical cadence. Should this be the case, the chordal toggling *after* the weaker [AVC] becomes part of a codetta rather than an R[P] passage leading into the [P-AVC] marked by the Disney version. Perhaps the

content of the WB, the focus on live action orchestra, allowed less precise mickey-mousing than the fully animated *Flamingos* version.

Figure 3.3 demonstrates the extensive use of rhythmic montage to support the final cadence in "Poor Unfortunate Souls" from *The Little Mermaid* (1989).[92] As discussed in Chapter 2, the use of rhythmic montage that the R[x]-[y] sync adapts for analysis pushes the audiovisual action forward. While the [AVC] cadences of Figures 3.1 and 3.2 happen at a steadier rate, Figure 3.3's cadences happen at a quicker, less periodic rate and combine use of shot changes, character, and camera motion. In this [AVC] complex, a V-i cadence underlies the lyrics "This poor unfortunate soul!" The G# to C# harmonies punctuate the final [I-AVC] and lead into a closing chromatic descent over alternating scale degrees 1 and 5 in the bass. The chromatic scalar descent in the orchestra music embellishes the underlying harmony before reasserting the tonic key at the end with V-I in the bass and Ursula singing a tied-over scale degree 1 in the voice line.

Shot changes synchronize to almost every beat of the passage leading into the final cadence. Each sync point represents an R[S]-[I^2] individually, but in the context of an [AVC] complex, the initial sync point would point towards an R[P]-[I^2] whereby the [AVC] occurs at Ursula's final word "soul." The sudden increase of sync points on the rhythm line over a VI^7-V-i progression unapologetically calls attention to the audiovisual interactions as the music leads into a prolonged V-i codetta. The correlation here focuses viewers on recurring sync points between harmony and shot changes, with less emphasis on the melodic aspects. It is as if dramatic effect

[92] Eisenstein 1977, 74. Eisenstein identifies his rhythmic montage as shot changes that go beyond the "mathematically determined length[s]" to create rhythmic violations and events away from the bar line. Adapting this idea, what I consider rhythmic AV cadences synchronize the final chord or R[P] passage with marked musical and visual acceleration, backed by cadential action as defined by music-theoretical conventions. This involves traditional V-I motion and any of the cadence types such as PAC, IAC, and HC.

For media reference, see: Disney. "The Little Mermaid | Poor Unfortunate Souls | Disney Sing-Along," YouTube video, 4:11, https://youtu.be/Gi58pN8W3hY?si=5jpwMKmh2flgzruR.

achieved through Ursula's choppy lines is prioritized over melodic coherence, relying on frequency of occurrence, sync correlation, and formal function to define the [AVC].

The final cadence of "Poor Unfortunate Souls" in Ursula's vocal part ends on an imperfect authentic cadence with scale degree 5 in the voice. Visually, the narrative is also imperfect in a sense that Ursula's final sync point cuts to Ariel and not Ursula. Ariel signs her name at the final sung word "soul," and the following character motion then cuts to the signed scroll at the fourth beat of that same measure. Only at the orchestral cadence in the last measure does the camera pan to Ursula (a substitution marked by the gray dot) followed by a final cut. It is almost as if, the initial [AVC], being imperfect, is corrected by the final two visual accents on Ursula, the main voice of "Poor Unfortunate Souls." This is further reinforced in the melodic line, where scale degree 5 also moves to scale degree 1.

It is possible for an [AVC] and its rhythmic shot changes to coincide with a form of phrase extension along with William Rothstein's four-bar groupings that reflect *phrase rhythm*.[93] This is strongly evident in "Under the Sea" from the same film.[94] Seen in Figure 3.4, Sebastian's [AVC] is more melodically driven than Ursula's example. His motivic repetition in mm. 1–6 of the Figure 3.4 transcription reveals a forced consistency of visual techniques recurring on E4, the highest note of the motive. Generally, something happens every time E4 is sounded in mm. 1–4, whether it is a synced shot change marked by the red arrows or a camera motion substitution marked by a gray dot. The side arrows similarly mark a presence of camera motion, most often a form of pan or zoom.

---

[93] Rothstein, 22–25. Since R[P] indicates some phrasal relation, *phrase rhythm* can equally apply to an [AVC], where the cadence is part of a climactic phrase that leads to the cadential point and resolution. As a reminder, phrase rhythm groupings are also decided by music-theoretical tonal and durational conventions such as melody or harmony.

[94] A YouTube version can be found here: DisneyMusicVEVO, "The Little Mermaid - Under the Sea (from The Little Mermaid) (Official Video)," YouTube video, 3:01, https://youtu.be/GC_mV1IpjWA?si=fgue5Wp2HbEi5RKl.

In Figure 3.4, the basic closing V-I harmony occurs at the [P-AVC] annotation in m. 8. The [AVC] can also set off phrase extensions in areas like mm. 8–11, where a codetta in the musical accompaniment occurs after the supposed cadence on the downbeat of measure 8. Like the orchestral codetta in "Poor Unfortunate Souls," the sync points in the "Under the Sea" codetta switch from mirroring the vocals to the musical line that elides with the cadence in mm. 7–8. Sync points in mm. 8–11 rapidly increase, accenting almost every note. This audiovisually emphasizes a phrase extension that William Rothstein terms a *suffix*, or the extension of a "basic phrase's closing harmony."[95] The sudden increase of visual accentuation contrasts the preceding R[P]-[AVC] complex that features constant camera motion and non-periodic shot changes, further dynamizing the sequence and differentiating between the bracketed [AVC] complex and the bracketed "instrumental [AVC]."

Figure 3.4 exemplifies how synchronization can switch tracks between sung and instrumental parts to include a total tally of all the sync points. Doing so demonstrates the generalizing power of the sync profile, as it can map the rhythmically oriented visual accents onto musical accents. In this case, the increase of surface rhythms moving beyond the vocal line is visually synchronized to help energize, correct, and reinforce the final [AVC] of "Under the Sea." This generalized sync rhythm line in Figure 3.4 also marks an important facet of the [AVC]: the ability to "correct" or "reinforce" a cadential moment. One reading of this passage would end the [AVC] on the downbeat of the excerpt's m. 8. A reading of this "ending" would emphasize the downbeat visual accent that punctuates the end mm. 7–8. However, the succeeding passage features an alarming increase in sync point frequencies. The instrumental part picks up on this and reinforces the ending on C5 through the final cadence in m. 11 (also

[95] Rothstein, 70–73

over a V-I harmony). As if to correct the lack of marked visual accentuation correlating to Sebastian's voice line (mm. 6–7), what would be the codetta picks up the V-I in m. 11 with a change on each harmony, clarifying the more ambiguous preceding visual narrative.

The presence of camera motion and absence of shot changes can problematize analysis if I adhere to Eisenstein's original analytical form, matching shot changes to an action at or away from the musical bar line.[96] However, incorporating camera motion as substitutes of shot changes responds to technological changes adopted by Disney and other animators. A clear example of how camera motion can be used as the main visual accent and replace shot changes in an [AVC] is shown in Figure 3.5. Here, Figure 3.5 excerpts the last cadential passage in "Honor to Us All" from Disney's *Mulan* (1998). Camera pans visually punctuate each female character who sings "please bring honor to us all." Figure 3.5 is recognized as an [AVC] because it coincides with accelerated harmonic and visual motion enough that missing cuts are understood to be missing. In this figure, the grayed arrows indicate the camera motion substitutions where shot changes would be expected, considering the examples set forth by prior Figures 3.1–3.4.

Character motion can substitute for shot changes in the same fashion as camera motion. The excerpt from "Hellfire" (*Hunchback of Notre Dame*, 1996) in Figure 3.6 exemplifies how camera and character motion may punctuate parts of an [AVC] complex in lieu of shot changes.[97] Figure 3.6 features all three forms of visual accents and accelerated harmonies with increased visual rhythms, but the cadence of Frollo's vocal part in the fifth measure of the excerpt and the

[96] The "Battle on the Ice Cue" graph, published in 1943, existed in a time before technological advances allowed incorporation of extensive camera motion and character animation techniques.
[97] Media reference: Walt Disney Feature Animation, *The Hunchback of Notre Dame*, directed by Gary Trousdale and Kirk Wise, Walt Disney Studios, 1996, streaming video, 1:37:00, accessed July 2024, Disney Plus, https://www.disneyplus.com/movies/the-hunchback-of-notre-dame/796BIJb9AlES.

orchestral cadence in the last two measures are synced to simultaneous camera and character motion and no shot changes.

In Figure 3.6's ending of "Hellfire," Frollo's vocal line closes with irregular phrase length, mainly because of the augmentation of his melody in mm. 2–5. There are only two shot changes in the passage, one starting the augmentation and one in the codetta's cadence. Like "Under the Sea," the visual component switches to the instrumental and choral part, separate from Frollo's vocal part that ends in m. 5 of the excerpt. Camera motion averts the use of shot change at the harmonic V-I of mm. 5–6. The switch over to the choral part moves the final shot change to the V-I motion in m. 9, reinforced by camera motion (zoom out, marked by light gray dot) in m. 10 with the lower octave reiteration of Eb in the orchestra part. Frollo's poses and movements, marked by sync point arrows in the figure, rhythmically correlate to the downbeat of each measure, enforcing Frollo's final three notes paired with lyrics "she will burn" (mm. 3–5 of excerpt). Understanding onscreen motion as substitutive of shot changes, the visual accents exemplify markedness that allows interpreting sync point analysis on [AVC] features such as harmonic closure through the scale degree 5 to 1 in mm 4–5, or the octave reiteration of scale degree 1 in m. 9 of the excerpt. The result is dynamic motion that directs attention to the cadences while subverting traditional shot changes.

As a final example of [AVC], Figure 3.7 presents the final climactic passage found in "Zero to Hero" from *Hercules* (1997). In this example, the downbeat shot changes in mm. 1, 4, and 6 accelerate into an AV cadence m. 12 of the excerpt. The music switches synchronization from the vocal to the instrumental section, leading to visual activity on every sounding note, some of it substituted by character motion/poses (dotted side arrows) or camera motion (solid side arrows) and are also highlighted in Figure 3.7's annotations. The voice line does not feature

a clear cadence until the very end, when the muses' melismatic performance ambiguates precise pitches. Therefore, sync to the accompaniment's chromatic ascent provides more solid, recognizable points of reference in an overload of audiovisual senses.

In the seven examples presented so far, the [AVC]s are all final cadences that exhibit recurring principles.[98] First, the focus on film rhythm ignores non-musical audio such as dialogue, voice-over narration, and sound effects (unless the sound effect is an R[S] sync written in the score). The audio component is strictly based on film music bears its own meanings and implications. This is a necessary simplification to focus analysis on the inner workings of music in relation to image.[99] As such, the second principle is that there is no differentiation between diegetic, music heard in the film world, and non-diegetic music, sounds heard outside the film world. This generalization is made in favor of studying the immediate effects of synchronous audiovisual rhythm.

Table 3.1 summarizes the table of sync points with the addition of the [AVC] subset. The [AVC] is therefore dependent on the definitions of [music accent] and [visual accent] as any of the other sync type categories. As a result, visual techniques are simplified and restricted to Eisenstein's definition of rhythmic editing, camera motion (most often zoom and pan), and character motion and poses.

The third principle for [AVC] cadences is that the sync points mirror Rothstein's idea of cadential closure preempted by "drive" or "marked acceleration" to the [AVC]. In Figures 3.3–3.7, there is increased [visual accent] activity that drives the visual action to the [AVC] point.

---

[98] These principles draw from the four analytical generalizations based on Eisenstein's graphic analysis and listed in the Literature Review section.

[99] Eisenstein's graphic analysis, he does not consider non-musical audio because the "Battle on Ice" cue's primary sound source is Prokofiev's score. There is no speech or sound effects until after the cue. I base my analysis on the fact that Eisenstein's technological limitations produced a scene that includes only music and image, an association that forefronts the standalone songs and music videos that serve as case studies in this dissertation.

This is especially true for “Poor Unfortunate Souls,” “Under the Sea,” and “Zero to Hero,” where a rush of shot changes accompanies a musical cadence with V-I harmonies. In the case of “Honor to Us All” and “Hellfire,” substitutive camera and character motion quicken onscreen action at the cadential points, synchronizing markedness or dramatic gestures to the cadential notes. In “Hellfire,” viewers are presented this “acceleration” twice, first with the cadence in Frollo’s vocal part and second with the orchestra’s final cadence.

In many cases, animated standalone songs or music videos may not have or require transcription to pinpoint an [AVC]. Protonotation or dot notation can be used to analyze [AVC] without a musical transcription.[100] Figure 3.8 illustrates how dot notation generalizes both the musical surface and visual events as happening on a structural beat of the indicated measure and time signature, disregarding the lowest-level durations unless necessary.[101] This means a visual event is understood to synchronize with beat 1 instead of a 32nd note of beat 1. However, the generalized notation still reveals the same principles of analyzing non-diegetic music for R[x]-[y] sync relations to [AVC] complexes that feature marked visual acceleration in tandem with musical closure.

This form of analysis is especially useful for rapped musical passages such as the “Cinema” Japanese animated music video from the *Hatsune Miku: Colorful Stage!* (also known as *Project Sekai: Colorful Stage!* in Japan).[102] This rapped excerpt does not feature the more

---

[100] This methodology is reminiscent of Thomas and Johnston’s discussion of the use of bar sheets (a.k.a “dope sheets”) in Disney animation, where blank blocks generated from musical measures were used to plan and even “mickey-mouse” music to visual action. However, the generalizations imply a reversal of the process, having analysis reverse-engineer sound points from visual accents. For more, see Thomas and Johnston, 288–89. Details such as dialogue and sound effects do not precede the selection of music or metronome beats set by the director and musician, something that Eisenstein also did for *Alexander Nevsky*.

[101] Fred Lerdahl and Ray S. Jackendoff, *A Generative Theory of Tonal Music* (Cambridge: MIT Press, 1983). The dot graphic notation reflects Lerdahl and Jackendoff’s dot notation examples and the well-formedness rule that my illustrated metrical level consists of “equally spaced beats.”

[102] Media reference: プロジェクトセカイ カラフルステージ! feat. 初音ミク. シネマ / Vivid BAD SQUAD × KAITO, YouTube Video, 1:19, https://youtu.be/8m1cUgLepmY?si=GmO1evPFbLgF4A18.

flexible rubato found in Ursula's performance or Frollo's dramatic declaration but presents 3D and 2D animations rigidly synced to camera motion and editing rhythms. My phrase groupings are determined by observing Lerdahl and Jackendoff's second metrical preference rule (MPR) that "weakly prefers" an early metrical anchor in a group. The rigid tempo of the music is interpreted at approximately 125 BPM. However, according to Justin London's theories of rhythm and meter, faster or slower BPM at division of 63 BPM (rounded up from 62.5 BPM) or 250 BPM are possible to an individual listener. According to London, meter is a form of entrainment, and I argue that the figure's visual rhythms function at 125 BPM, informing an 8-bar phrase.[103]

The [AVC] in Figure 3.8 ends an 8-bar phrase running at approximately 125 beats per minute (BPM), cadencing before moving into the chorus section.[104] Each bar grouping in Figure 3.8 is indicated by a numbered, solid black dot marking the downbeat, generalizing four beats with the remaining gray dots instead of attending to each rapid note in of the pop rap. With the buildup of irregular sync points in groupings 6 and 7, the [AVC] is markedly apparent in grouping 8, where an accelerated presence of flashing shot changes rhythmically signifies and reinforces the phrase end.[105] This demonstrates how acceleration and steady entrainment of shot changes reinforce the [AVC]. Even without clear indication of the underlying harmony, visual

[103] Justin London, *Hearing In Time: Psychological Aspects of Musical Meter* (New York: Oxford University Press, 2012). It is possible that a listener entrains to three different levels of meter, depending on the tactus. Other stimuli such as the drum track, the vocal speed, or anything else may compel the listener to one BPM over the other. Most of the time, the tactus is the notated beat, but in this case, it is a retrospective grouping of the beats strain.
[104] BPM approximations are retrieved by (1) utilizing a tempo tapping program that determines the beat, and (2) processing the tracks through online BPM calculation software. For this study, songbpm.com and soundplate.com track analyzers were used, with only a 1-2 beat margin of error. Depending on the software/method, "Cinema" is analyzed as between 124-126 BPM.
[105] The musical interlude preceding this excerpt features a still frame with animated chalk overlays that are, in retrospect, animated to the 125 BPM tempo.

rhythm can mark phrase endings, earning equal importance in indicating [AVC], especially in this more raplike passage.

3.2: The Vococentric Shot Change

There is a recurring correlation between rhythmic shot changes and a sense of sonic fidelity, more specifically showing who or what is sounding/singing. This is typical for ensemble songs that include extensive choreographies or intertwining vocal parts. Disney musical numbers and music videos for animated performing groups are among the case studies with the most occurrences of these sonic fidelity-based shot changes. Since the onscreen action is related to non-rhythmic musical features such as timbre, pitch, and melodies, these [ICO] type visual accents can be considered a subset that is "vocal-centric," or focused on identifying the source of voice. While the [VCO] can easily be labeled [ICO] of which it is a subset, indicating a sync point as [VCO] refines the sync point function. Describing who or what is singing now not only provides another layer of specificity in analysis but also differentiates which sync points are simply iconic to the music and which are iconic of the vocal line. The specificity provided by the [VCO] subset aids in analyzing more intricate ensemble songs or pinpointing moments where characters speak through song. The following technical definition describes the [VCO], or "vocal-centric," relation:[106]

[VCO] $\subseteq$ **R[x]-[ICO]** sync point types

Where [x] = R[S], R[P], or R[D]

[106] This is what I consider vococentric in nature: a concept that is adapted into an animated musical format but still reflecting Michel Chion's concept of the vococentric cinema, or images influenced by the agency of the voice and dialogue.

The [VCO] is particularly important in Disney animated musicals where sung melodies drive characters' internal or story montage narratives. Many numbers may also feature two or more singers, necessitating visual accents that direct attention towards sound source and fidelity to indicate who is singing at a given moment. While [VCO] types can conflate with [ISO] types, the main distinction is that [VCO] clarifies who/what is singing or sounding while [ICO] clarifies how and what pitch or timbre is being represented. Therefore, the [VCO] is a subset of [ICO] relations that refine the specificity and function of an iconic visual accent. This label can be used to further specify the type of [ICO] relation. To indicate this, the label R[x]-[VCO] will be used in analysis. Table 3.2 adds the [VCO] subset to the table of sync tables.

Given the "vocal-centric" nature of Disney musical numbers, I adapt the term *vococentrism* to label any visual accents that not only punctuate music, but mark and synchronize with a singing voice. Since the Disney standalone songs often develop character and monologues through song and lyrics, the *vococentricism* is often presented through vocals, hence my interpretation of "vocal-centrism." Because the markedness is instantaneous and signals a change in voice or vocal quality, there is lesser chance of an R[D]-[VCO] relation because the [Duration] property mainly accounts for non-pitched synchronicities or sync points that, while latched onto a phrase or melody, only refer to the rhythmic properties of the music. In this case, an R[P]-[VCO] relation would imply duration, as the singing character needs to maintain an entire musical phrase on their own without any interspersed visual accents, camera movement, or shot changes.

A R[S]-[VCO] relation focuses on clarifying a change in character voice which affects timbre, but the change does not outline a musical phrase. The character sings a melodic snippet or motive but does not perform a complete musical phrase, which warrants the R[P]-[VCO]

label. Figure 3.9 serves as example of a R[S]-[VCO] relation from "Zero to Hero," where three different singers split a musical phrase into parts through visual accentuation interspersed with other sync point types. In a short passage that is visually and musically overwhelming, the R[S]-[VCO] sync points help clarify who is singing amidst the rapid pace of shot changes.

Returning to "Cinema" of the *Colorful Stage!* series, Figure 3.10 is an example of [VCO] visual accents that substitute for traditional shot changes. Again, the musical phrase is split among characters, and the zooming visual accent (camera motion marked by solid sideways arrow) marks one character after another. As none of the characters completes an entire musical phrase, the relation is R[S]-[VCO], where [sonic] refers to change in character voice and [VCO] specifies vococentricism as the reason for visual accentuation. Just as in the [AVC] example of the same song, the generalization of the rhythmic surface helps clarify how the camera zooms (marked by side-facing arrows) out every four beats, focusing on one of the five characters, set in an immobile, 2D key-pose. The stillness of characters set in key-pose demonstrates how camera motion can visually energize motion and accent onscreen "action," albeit simulated.

An R[P]-[VCO] relation is more encompassing than the R[S] relation. An example of this is Figure 3.11's "Love is an Open Door" from *Frozen* (2013). In the opening, both Anna and Hans sing an entire melodic phrase, with visual accents cutting to the singer at the beginning of each melody. With a focus on character and camera motion, there are fewer instances of abrupt shot changes that distract from the [VCO].

A case study of *Steamboat Willie* (1928) demonstrates the [VCO] functions in the fourth and fifth repetitions of the precomposed folksong "Turkey in the Straw."[107] This example also

[107] In the animated adaptation, an arrangement of "Turkey in the Straw" is repeated six times. The fourth and fifth repetitions feature the best instances of rhythmic sync points in the musical sequence. The other repetitions will be revisited in the next chapter for their more periodic or *metric* properties and function.

points towards early "vocal-centrism" in Disney's first sound-synced animation. In their attempt to create a serious narrative and believability in sound sources, certain timbres must be attributable to each animal's "voice," albeit with the absence of lyrics. In those two rotations, shots occur nonperiodically at the sync points in Figure 3.12a. Sync point 1 in Figure 3.12a features a close-up of Mickey pulling on the goose to focus on the bird's "voice" that parallels the musical melody. The shot change occurs on the weaker fourth beat and only reflects the duration of the musical note, therefore earning status as an R[D] type sync.[108] The focus on the goose's voice warrants an R[D]-[VCO] label to indicate that a durational shot change was made on the weak beat of the music (the "what") with focus on the sonic fidelity to the singing goose by means of a shot change (the "how"). This R[D]-[VCO] similarly applies to the shot change at the end of the repetition, where a cut to a group of nursing piglets (sync point 6) is placed into narrative focus.

In the fifth repetition, the shot changes to a perched parrot (sync point 3) and accomplishes the same R[D]-[VCO] function. There is a slight change to this trend at sync point 5, where the shot change emphasizes the transition from the goat phonograph to a cow that serves as a xylophone for Mickey. Although the shot change is another R[D] emphasis on the music, the visual function is [ICO] instead of [VCO]. It is so because the cow is not singing, but the xylophone sounds are represented by the cow's teeth. This differentiation emphasizes, in practice, the fine line between an [ICO] and [VCO] function, where the former stresses sonic representation while the latter stresses sonic fidelity and the "voice" of the immediate sounding character. In other words, although a general R[x]-[ICO] and R[x]-[VCO] may appear similar,

[108] The shot change occurs milliseconds between the end of the third and beginning of the fourth beat. As typical to hand-drawn animations and technological limitations that have yet to allow precise timing of shot changes, a generalization towards the fourth beat facilitates analysis. Even so, the location of the shot change

[ICO] is a generalization of how and what pitch or timbre is being *represented* and [VCO] answers the question of who/what character is *singing*, whether with lyrics or vocalizations. Figure 3.12b visually summarizes the shot changes for all repetitions.

3.3: Conclusions

The subsets [AVC] and [VCO] further refine the function of audiovisual sync points laid out in Chapter 2. By specifying the R[P]-[y] sync as an [AVC], analysis can draw immediate attention to cadential points that play a critical role in defining musical form and function. Pinpointing how visual accents embellish the musical aspects illustrate the audiovisual intertwining that sync points draw attention to. Identifying an [ICO] sync type as [VCO] forefronts a recurring practice where the visual accents direct attention to a change in singer, to ensure that viewers clearly know who is signing when the timbre, or sonic fidelity changes among ensemble singers or—in the case of *Steamboat Willie*—personified animal characters.

In the next chapter, these subsets and their rhythmic sync type sets are presented as analytical "blocks" that build a bottom-up view of *metric* montage, or what Eisenstein considers visual editing at consistent and mathematically determinable lengths. These resultant metric "supersets" present enough consistency to parallel with the musical concept of meter and hypermeter, projecting sync points and their narrative functions over larger time spans and at larger formal levels in terms of the musical composition and film runtime. As the R[x]-[y] sets and subsets demonstrate how sync points create local and immediate audiovisual narratives from point A to B, the upcoming *metric* supersets frame and provide more bird's eye views of sync point total functions.

Chapter 4

Metric Synchronicities

4.1: Background

In this chapter, the rhythmic sync point types and subsets presented in Chapters 2 and 3 serve as building blocks for two more sync point types: the metric and hypermetric. Though the terms *metric* and *hypermetric* invoke musical meaning, the two terms in this chapter refer to a (hyper)metric rhythm that is based on audiovisual relations that correspond to, but is independent from, musical hypermeter. Therefore, the metric and non-cadential synchronicities in this chapter are not determined solely by musical hypermeter nor visual hypermeter, but by the synchronous unions between the two. This chapter's use of the "metric" refers to Eisenstein's definition of metric montage as visual editing at consistent and mathematically determinable lengths.[109] He even states that metric montage is useful for organization as well as rhythmically connecting film to audience.[110] I map Eisenstein's definition onto my analytical framework to open possibilities for identifying sync points positioned at equal distances and durations. To this end, this chapter seeks to demonstrate how sync points of equidistantly determinable lengths serve an organizational and narrative-framing function in the film rhythmic narrative.

Figure 4.1 revisits "Cinema" from the recent *Hatsune Miku: Colorful Stage!* series to provide a preliminary example of metric synchronicity. This *anime* example from prior chapters features a collection of rhythmic sync points that exhibit periodicity and are spaced apart evenly

[109] Eisenstein, *Film Form*, 72–73. Eisenstein wrote that "The fundamental criterion for this construction is the *absolute lengths* of the pieces. The pieces are joined together according to their lengths, in a formula-scheme corresponding to a measure of music. Realization is in the repetition of these 'measures.'" He later writes that metric montage is recognized "Not by *impression* as perceived, but by *measurement*."

[110] Ibid., 73. In the original text, Eisenstein states that "Its [metric montage's] clarity can bring into unison the 'pulsing' of the film and the 'pulsing' of the audience. Without such a unison (obtainable by many means) there can be no contact between the two." Brackets, mine.

for spans of two or more notes. This type of sync marks metrically important areas of music such as the downbeat and provides an effect of consistency and organization. Therefore, they are elevated to a metric status, labelled "M," since the sync points occur on every equidistant downbeat six times in a row. There are no other sync points in between to disrupt the even spacing. This supports Eisenstein's metric montage definition as visual editing at consistent and "mathematically determinable" lengths. Because of these mathematically determinable lengths, sync points can still be pinpointed in unannotated scores such as "Cinema's" raplike sequence in Figure 4.1.

A sync point can be hypermetric if a collection of metric sync points spans many measures of music and accents parts of phrase structure, per William Rothstein's *phrase rhythm* principles.[111] The effect of this sync type is looser than rhythmic and metric in terms of immediacy, but it organizes and frames melodic and phrasal structure more clearly than the downbeat-oriented metric sync. An example of this is Figure 4.2, a revisiting of Disney's *Flamingos* where hypermetric sync points form over a collection of metric sync points that coincides with the passage's phrase structure. The category "HM" is formed from the M-sync in Figure 4.2 and emphasizes the downbeat of the scalar phrase while adhering to music theorists Lerdahl and Jackendoff's well-formedness rules for preferring the downbeat for determining grouping structures, i.e., phrase structure in this case.[112]

---

[111] William Rothstein, *Phrase Rhythm in Tonal Music* (New York: Schirmer Books, 1989). Rothstein was influenced by his predecessors, Carl Schachter, Heinrich Koch, and other 18th century theorists who favored Schenkerian methodologies and 4-bar phrases. Any irregular phrase lengths, like 5-bar phrases require a degree of reinterpretation or the acknowledgement of phrase expansions. My application of hypermeter to identifying sync points in modern film music pushes boundaries of strict Schenker-based phrasal analysis.

For additional information on hypermeter and irregular phrase groupings, see Brian Edward Jarvis and John Peterson, "Alternative Paths, Phrase Expansion, and the Music of Felix Mendelssohn," in *Music Theory Spectrum* 41, No. 2 (2019): 187–217, https://doi.org/10.1093/mts/mtz009.

[112] Fred Lerdahl and Ray S. Jackendoff, *A Generative Theory of Tonal Music* (Cambridge: MIT Press, 1983). I believe that application of the well-formedness rules, based on how expert listeners make sense of musical stimuli, suit the esthesic aspects of film music analysis and the focus of this dissertation.

Recall that the analytical framework presented in Chapter 2, n[x]-[y], is a formulaic representation of an audiovisual sync point. Like rhythmic types, the *n* is a placeholder for the metric ("M") or hypermetric "HM" *rate* of synchronicity. Variable [x] remains a definer of musical accent type, while [y] represent visual technique. Since rhythmic sync types are building blocks of (hyper)metric types, the latter become sync point *superstructures* made of the former. As superstructures, (hyper)metric sync functions to organize, frame, and encompass larger stretches of audiovisual action. As a result, (hyper)metric sync creates comprehensible formal pockets within an ongoing musical "auditory stream." They are instrumental in differentiating correlations to musical form from what are otherwise ambiguated collections of rhythmic sync points. This is possible because (hyper)metric sync is consistent, equidistant, periodic, and regulated.[113]

Metric sync points prioritize consistent time points through calculable durations that parallel Justin London's concept of interonset intervals. These intervals, or IOI, are defined by beats per measure/minute in relation to musical phrasing. Eisenstein's "measurable" visual rhythm correlates with Justin London's interonset intervals, or the time spans in milliseconds or BPM between two musical articulations. The IOI is flexible in analysis because it considers the space between articulations, not necessarily durations. It also allows scoreless analysis as long as sync point analysis sets parameters for IOI equidistance as *consistent durational values among three or more sync points*. Figure 4.3 revisits "Cinema" from the *Colorful Stage!* series.[114] With a

[113] These characteristics support McAdams and Bregman's notion of "internal consistency" within a "whole" auditory stream. While this idea originally pertained to audio sound and production, the idea's essence is compatible with my framework that treats audiovisual filmic experiences as a linear process accompanied by an ongoing auditory stream. In my theoretical framework, the filmic space (per Pudovkin) is made up of intertwined auditory and visual (technique) streams that are punctuated by sync points n[x]-[y]. For the original source, see Stephen McAdams and Albert Bregman, "Hearing Musical Streams," in Computer Music Journal 3, no. 4 (1979): 26–60, http://www.jstor.org/stable/4617866.

[114] The [AVC] of this excerpt is discussed in Chapter 2.

constant BPM in the background, what seems to be R[S]-[VCO] focused on timbral changes of the singers are consistent enough to form a stream of metric sync. Therefore, the labels transform from the initial R[S]-[VCO] to M[S]-[VCO].

My definitions are based on Rothstein's music-theoretical principles, and the (hyper)meter is expressible through Schenkerian terminology, from which Rothstein's writings heavily develop. While Schenkerian analysis is not directly applicable to the audiovisual framework, there are three underlying concepts that reinforce the triangular relationships among this dissertation's sync types. In other words, rhythmic, metric, and hypermetric syncs can be paralleled to the three Schenkerian concepts of analytical "levels," or strata:[115]

- R-sync concerns the "foreground," or all the notes in a score/transcription.
- M-sync concerns the "middle-ground," or the emergent larger patterns to which "all" notes are subordinate to.
- HM-sync concerns the "background," or visually accented and deeper large-scales structures and form.

At its core, Schenkerian analysis is inherently arrhythmic by the time a background graph is produced. However, its concept of "layers" can apply to sync point analysis even if the sync analysis framework is rhythmically and durationally oriented. Rhythmic, metric, and hypermetric sync points correlate rates of visual accentuation to musical levels in a similar manner. This is possible due to the framework's treatment of the film as a product and hypertext of audiovisual

[115] Though Schenker's concept of the fundamental structure is considerably "arrhythmic," to the extent that background graphs do not need exact rhythmic values, my parallels are made to point to the level of detail sync point types concern with. Film music has had prior tonal-based analytical frameworks such as the more Schenkerian approaches (Cochran 1986) or tonal design narratives (Rodman 1998). For more, see Alfred Cochran, "Style, Structure, and Tonal Organization in the Early Film Scores of Aaron Copland," Ph.D. diss., (The Catholic University of America, 1986). See also: Ronald Rodman, "There's No Place Like Home": Tonal Closure and Design in The Wizard of Oz," in *Indiana Theory Review* 19 (1998): 125-43. http://www.jstor.org/stable/24044541.

interactions created from possibilities in a filmic space, per Alexander Pudovkin.[116] In this chapter, R-, M-, and HM- sync types correlate to fore-, middle-, and background level of details. However, the framework turns these foundational philosophies on its heels: instead of analyzing how the background controls the presentation of the foreground, sync point analysis observes the "foreground" for emergent relationships and "builds" meanings across three levels of synchronicities.

4.2: Terminology and Definitions

Metric sync is an umbrella category that includes various subsets: metric sync as a standalone unit, [AVC] passages, and hypermetric sync subsets. Hereafter referred to as (hyper)metric, all metric sync types are measurable because they mirror durational values of musical notes or measures. For example, a metric sync point must span two or more *equidistant* beats of music, or a hypermetric sync spans two or more musical measures. Therefore, (hyper)metric synchronicities utilize duration to create consistency and periodicity while calling attention to large-scale harmonic and/or phrasal features that span many equidistant intervals (beats/measures) of music. In technical terms, a metric sync point n[x]-[y] is defined as:

M • [x]-[y] = M[x]-[y]

- Where "M" = Metric sync
- And [x] = music accents [S], [P], and [D]
- And [y] = visual techniques [ISO], [ICO], [VCO], or [I^2]

[116] Vsevolod, Pudovkin, *Film Technique and Film Acting* (New York: Lear, 1949).

Since metric sync is durationally defined [D], it can still emphasize the duration of a sonic [S] or phrasal [P] music accent [x]. These music accents are consistently marked by any of the visual techniques (i.e., shot changes, character motion, and camera motion) that are isographic [ISO], iconic [ICO], vococentric [VCO], and combined isographic-iconic [I^2]. Finally, to differentiate the *rate* of synchronicity, the M- and HM- levels must be defined as follows:

- Metric sync points operate on the beat level, consistently encompassing two or more beats between each marked sync point.
- Hypermetric sync points operate on the measure level, consistently encompassing two or more measures of music between each marked sync point.

The further clarify this, the hypermetric labelling formula defined as follows:

HM • [x]-[y] = HM[x]-[y]

- Where "HM" = Hypermetric sync spanning two or more measures
- And [x] = music accents [S], [P], and [D]
- And [y] = visual techniques [ISO], [ICO], [VCO], or [I^2]

The idea of hypermetric sync transcends the original, "metric," definitions in Eisenstein's writings. Metric sync is more marked than hypermetric sync, in that it is easier to perceive shorter spans of audiovisual data and closer to the immediacy of the filmic experience. However, I believe that conceptualizing an additional category for hypermetric sync facilitates identifying higher degrees of regularity in larger spans of audiovisual action. Tracking higher degrees of

regularity at the HM[x]-[y] level allows analysis of formal delineations, macro-level sync point functions, and broader narrative implications.

Figure 4.4 presents a basic application of the M[x]-[y] sync point label. Here, the M[P]-[ISO] sync points are rather straightforward in definition. Figure 4.4 cues in on the "Turkey in the Straw" scene from *Steamboat Willie*, with focus on the first three repetitions. Two visual accents per repetition are annotated in the figure and delineate the A and B sections of the precomposed song. The M[P]-[ISO] would initially be recognized as a rhythmic sync point. However, a retrospective realization that the metric sync points are *equally spaced* in terms of duration *elevates* the rhythmic sync point to a metric status. Over the course of three repetitions, a larger organization pattern emerges, and a 16-measure hypermetric relationship emerges. The upwards arrow in the annotation represents the emergent hypermetric HM[P]-[ISO] sync type *built up* (e.g., *elevated*) from the pre-existing metric sync point. This simple passage sets up five principles of (hyper)metric sync:

1. Metric sync points must have consistently *equal* durations among occurrences, encompass at least *two or more* beats in between, and can encompass a span of measures regardless of phrase structure.
2. Metric sync points must accent at minimum three equidistant occurrences; if not, they maintain status as lower-level, rhythmic sync points.[117]
3. M[x]-[y] sync points that span two or more measures can *elevate* to *hypermetric* status "HM[x]-[y]," even in an absence of phrasal units.

[117] As explained later in this chapter, the tripartite rule adapts Hasty's and Mirka's metric projection theories. In this case, the three equidistant occurrences count the repetitions.

4. Since HM[x]-[y] points are built from M[x]-[y] sync points, they can elevate to HM status without meeting the minimum tripartite requirement if there are at minimum two consecutive M units.
5. (Hyper)metric sync points may alter an R[x]-[y] sync's original [x] or [y] variables and are not required to retain the original [x] or [y] qualities.[118]

Figure 4.4's phrase structure is clear, and the shot changes coincide exactly with the absolute durations of each eight-measure span A and B. Being more than four measures in length, the metric sync point is elevated to a hypermetric one, thus the HM[P]-[ISO] label. While one repetition of the sixteen bars features only two metric sync points, the entire experience reinforces a hypermetric reading by involving three complete repeats. This provides the preferred minimum to reinforce an HM[P]-[ISO] reading.

Analysis of M- and HM- sync types depict an emergent, large-scale narrative for the "Turkey in the Straw" sequence. Shown in Figure 4.5 are two differing streams of metric and non-metric sync types. Figure 4.5 visualizes the consistent visual accent rhythms for variable [y] in repetitions 1–3 as well as the unpredictable accents in repetitions 4–6. The hypermetric organization established in repetitions 1–3 creates an *expectation* for continuation in repetitions 4–6. However, the latter deviates from said expectation and features rhythmic shot changes in unpredicted areas. An ironic and humorous narrative arrives from the deviation from expectations, creating an audiovisual and esthetic reflection of Mickey Mouse's character. It sums up his animated deviation from the steamboat job to join Minnie Mouse in enacting a musical performance.

[118] For example, hypermetric sync points may generalize the specificity of onscreen action, emphasizing [ISO] over [ICO] relations.

In addition to the large-scale occurrences of Figure 4.4 and 4.5, M[x]-[y] sync relations can also occur in smaller passages and durational intervals. Due to the criteria that M- sync types can be built from consistent units of two or more measures, M[P]-[y] sync types can delineate cadential points built on [AVC] subsets. Figure 4.6 presents the developments of cadential metric sync in Figure 4.6a's *Moana* (2016) and Figure 4.6b's *Mulan* (1998). Here, annotated *rhythmic* sync points occur at consistent periods of "absolute lengths" and at the downbeat of every measure that complies with Lerdahl and Jackendoff's well-formedness rules for determining grouping structures.[119] These two examples illustrate how periodically recurring R[S]-[y] sync types become the building blocks that achieves M[x]-[y] status through consistent and equal-durational occurrences. Figure 4.6c provides an example of an R[P]-[AVC] complex that *does not* elevate to metric status. This is due to the varied periodicities among the sync points. Despite being all equal quarter note durations, the frequency of occurrence does not comply with Principle #1, that there must be two or more beats in between metric sync points.

In accordance with (hyper)metric Principle #5, the elevation from R- to M- sync type in Figure 4.6b features a transformation in [x] music accent type due to the shift in materials highlighted. As a result, the collection of rhythmic sync points that displayed *periodic consistency* (i.e., "absolute lengths") is *elevated* to a metric level, thus creating an M[P]-[I^2] sync point for the first iteration and M[S]-[I^2] for the subsequent ones. The first M[P]-[I^2] defines the beginning of the phrase, while the following M[S]-[I^2] reinforces the phrase structure through consistent frequency. The passage is not hypermetric since the sync point occurrences do not

[119] Lerdahl and Jackendoff, pp. 69–96.

span two or more measures. The [AVC] complex from the rhythmic iteration carries over to the metric level and consequently forms an M[P]-[AVC] complex.[120]

Recall that the [AVC] itself is a subset of the R[P]-[y] rhythmic sync type. The [AVC] subset can still exist at the metric level, as the label indicates that metric sync subsumes the [AVC] subset within the *superstructure*. Including the [AVC] label for the complex or cadence itself specifies the punctuating function of the metric sync. Since metric sync is built from the rhythmic version of a passage, it also indicates that the [AVC] cadence or complex is contained within a span of measures that fulfill Principle #2's requirement for a tripartite unit of equidistant sync points.

Figure 4.7 presents M[D]-[ISO] sync points without highlighting a phrasal structure. Built on originally R[S] sync points, the metric quality of the M[D]-[ISO] is based on Principle #1, or the "consistently *equal* durations among occurrences" that "encompass at least *two or more* beats" in between. Figure 4.7 revisits a familiar passage from Disney's *Flamingos* (*Fantasia 2000*), where a collection of eight R[S]-[ICO] occurrences transform into M[D]-[ISO] whilst fulfilling Principles 1, 2, and 5. This fulfills elevation to metric even if phrase structure is not outlined by the M[D]-[ISO] sync points that occur every measure.

This excerpt is also another example of rhythmic to metric sync elevations that do not maintain the same [x] and [y] values. The change is a result of a shift in audiovisual accentuation. On the R[S]-[ICO] level, the yoyo flamingo's choreography is iconic of the scalar passages. On the metric level, the M[D]-[ISO] indicates that the flamingo's choreography has a

[120] As a reminder, Chapter 3 states that an [AVC] complex includes a phrasal unit [P] that includes music that builds to and from an [AVC] cadence. The M[P]-[AVC] complex provides another form of [AVC] cadential action. It does not encompass [AVC] such as the "Zero to Hero examples in Chapter 3 because, despite technically having almost equidistant durations with sync points on every beat, the closeness and quickness of actions disrupt instead of establishing a metric constant. Such [AVC] cadences remain R[x]-[AVC] type.

visually marked change every measure of equal durations. These two annotated levels reveal the layers of detail in Disney's audiovisual narrative: that yoyo flamingo's actions are mirrored on multiple levels, through iconic representation at the rhythmic level and durational values at the metric level.

Figure 4.7's detail in mickey-mousing extends even further to the hypermetric level, where the HM[P]-[ISO] label emphasizes the hypermetric organization of the scalar passage. This emphasis on the [P] music accent and [ISO] visual technique variables indicates that [ICO] accentuation at the rhythmic level is generalized in favor of large-scale organization. In its entirety, all annotated levels in Figure 4.7 best demonstrate the sync point framework's utility from the local to larger-scale analysis. Analysis at the rhythmic level details how closely the yoyo flamingo's choreography mirrors the scalar passage in motion and velocity. Study of the metric level reveals that the rate of audiovisual change coincides with the mathematically determinable durations of each musical measure. Understanding the hypermetric level forefronts organization of larger musical spans that aid in comprehending stretches of audiovisual action according to phrase rhythm or patterned scalar passages as that in Figure 4.7. The nested simultaneity of all three sync rates makes Figure 4.7 an epitome example of extreme mickey-mousing.

All three levels of sync rates R-, M-, and HM- of Figure 4.7 provide tripartite support for a metric reading. The high level of audiovisual correlation is summarized as a pyramidic relationship in Figure 4.8, whereby the rhythmic sync types covered in prior chapters serve as building blocks for the higher metric and hypermetric levels. As a result, the bottom-up pyramid demonstrates an elevation from most-specific to most-generally detailed types of synchronicities based on identifying sync rates.

4.3: Metric Sync as (Mostly) Tripartite Process

Metric sync analysis is idiosyncratic because it straddles between retrospective and active realization. The goal of sync point analysis is to examine the dynamic sequence of audiovisual events that leads from one point to another, with the larger purpose of defining the role of metric projection in framing and organizing sync point functions. Therefore, this section combines sync point analysis with tripartite metric projection to describe the connective processes between sync point events. The resultant process is almost a reverse-engineering or simulation of the film reception process, technically explained through sync analysis.

Sync points are active because they unfold in filmic time and space as presented to viewers. As the film is audiovisually experienced, received information is processed for retrospective understanding of filmic content including, but not limited to, audiovisual information. In all preceding examples, metric sync must occur according to Principle #2 and in units featuring intervals of at least three or more notes. In other words, M[x]-[y] ≥ 3 equidistant sonic attacks or "tones" that provide a sense of stability in contrast to disruptive rhythmic sync.

M[x]-[y] type sync is characteristically retrospective because sync point elevations are recognized after the fact and are often facilitated with a musical transcription or score. As a result, (hyper)metric sync that is treated passively does not accurately reconstruct the dynamic film experience nor the hyper-manufactured, hyperreal "experience" created by an animated scene. However, (hyper)metric sync can take on a more active role if it is placed into a tripartite metric projection process.

Christopher Hasty's metric projection process is a music-theoretical framework that enlivens Neumann's projection theory and is in turn invigorated by Danuta Mirka's dynamic

version.[121] Hasty's 1997 theories of rhythmic and metric projection present a singular sounding tone as a point of potential growth.

Figure 4.9 is my visual summary of Hasty's theory in abstract space.[122] Tone #1, or the "potential" tone, is the starting point for any tripartite relation. It is a springboard, a point zero, that creates an expectation for Tones #2 and #3 to follow. It takes at minimum two tones (#1 and #2) to determine the *duration* of the potential tone and the subsequent. Usually, tones 1 and 2 are similar or identical in duration or content, creating an expectation that the third tone should be essentially *the same* as the first two (i.e. identically spaced in durational value). The duration informs Tone #3, which *determines* whether the tripartite relationship is confirmed or thwarted. For (hyper)metric sync, the tripartite relationship is strictly durational and does not need to involve other musical features. If Tone #3 is as identically spaced (in terms of duration) as Tones #1 and #2, then the tripartite structure *confirms* a unit of durational consistency. This means that any collection of rhythmic sync points can elevate to metric status. If the duration of Tone #3 differs from Tones #1 and #2, it thwarts expectations and creates a denial phase where Tone #3 becomes the new Tone #1 and repeats the potential-duration-determinate process. This projection model leads to two possible paths: one determinate (phrase-ending or unit-confirming) and one non-finite (requiring further projection). The cyclic orientation and reorientation to this cycle is

---

[121] A footnote for historical context: Twentieth-century theories grapple with these ideas an begin to enliven the viewpoints of energized music as motion, whereby the study of meter is dynamized as a process of temporal stimuli as opposed to dialectic products. This viewpoint is advanced by the formal studies of grouping metrical units, determining hierarchies, or identifying architectonic units in rhythmic music. These "energetic" attitudes shift the attention to cognitive and intellectual process to make sense of rhythm and meter.

Interactive projection is developed by Neumann (1959) and both theorized and transmitted by Hasty (1997). While Hasty also transmits ideas of Neumann (1959), Hasty unites ideas of rhythm and meter as a process, not just categorizations of connections between two points. In contrast, Neumann's original model separated meter and rhythm and claimed that they could be experienced/informed separately.

[122] Neumann's ideas are originally based in the "persistence/accumulation" phase (what would be Tone #3) are determinate of the tripartite structure. For instance, Tone #3 would inform what Tones #1 and #2 need to be. Hasty's angle reverses this idea and is based on the starting point, Tone #1, as the energetic springboard for Tones #2 and #3. This straightforward reversal is what makes Hasty's framework more active than Neumann's retrospective focus on Tone #3.

what enlivens the reception and projection of (hyper)metric sync points, bringing analysis closer to understanding real-time filmic processes.

Sync points can easily map onto the slots of Tones #1–3. From the filmic sense, Eisenstein's original theories were originally dialectic and retrospective in analysis. His approach was like Neumann's and less energetic than Hasty's framework. However, including Eisenstein's montage (film editing and shot cutting) techniques to the [y] variable of the n[x]-[y] formula allows the classic rhythmic cutting to join a palette of modern techniques (e.g., character animation and camera motion) and become part of the dynamic projection process. In this process, techniques such as montage shot changes encourage viewers to *anticipate* (e.g., project) sync points 2 and 3, which take the place of "tones" from the original graph. Figure 4.10 maps this tripartite projection trajectory onto the final cadence in "Cinema," from prior analyses. After a series of reaffirmations, the metric sync is thwarted by the [AVC] cadence that disrupts the metric projection train.

In music-theoretical scholarship, Danuta Mirka's application of the multiple parallel analysis system (2009) further dynamizes Hasty's theories. Mirka's theories derives meter directly from a composition, or the musical information plotted out by composers who can manipulate meter to higher complexities. My sync point analysis adapts Mirka's idea to claim that the audiovisual "meter" is in the filmic experience and hypertext. Sync points become a deliberated marker for audiovisual interactions, especially since animation is capable of being plotted, fabricated, and animated during film production. Figure 4.10's scoreless excerpt serves as an example of how metric sync is drawn from the musical hypertext itself.

Without annotations, Figure 4.10 would be a unidirectional audio stream. The annotations show that the metric nature of the music's repetitious pop rap and consistent rate of visual accent

techniques create an emergent and graspable sense of organization in the animated sequence. In Figure 4.10a, the red arrows indicate the normal projection of R[S]-[VCO] in real time. The arcing arrow annotated "tripartite unit" represents the group, *reinforming*, and elevation of R[S]-[VCO] to the M[S]-[VCO] level that is reinforced by subsequent M[S]-[VCO]. In Figure 4.10b, The M[S]-[VCO] units are remapped alongside the projection nodes to show that, while the functional labels change, the projection trajectory remains the same, thus preserving the overall rhythmic profile at an elevated level.

Mirka's theories imply that there are always two competing levels of interacting rhythms. I interpret that, for metric sync analysis, there are aways two competing levels of interacting synchronicities per given audio stream. This idea is much like how M[x]-[y] sync points projections always check against their R[x]-[y] counterparts to verify or deny metric status. Just as Mirka's theories exemplify, one level of rhythms will always be more salient than the other, though the roles may switch. Figure 4.11 illustrates a sync point trajectory that interweaves M[x]-[y] and R[x]-[y] relations in the opening excerpt of *Flamingos*.

The tremolos at the beginning of Figure 4.11 create a metric consistency over four measures. This meets the tripartite minimum for Principle #2 and includes the potential, duration, and fulfillment phases of metric sync projection. This projection is thwarted in measure 5. Although an M[D]-[I^2] occurs as expected on the downbeat, the second half note in m. 5 denies further projection of M[D]-[I^2] sync points and switches to an R[S]-[I^2] sync type. The "waking up" motif that appears as a rising bassline creates a new projection trajectory and maintains it across four equidistant sync points. The four R[S]-[I^2] sync points occur away from the "preferred" metric downbeat and contrast the metric consistency established by the "sun rays."

The new and offset projection trajectory erases the possibility for establishing M- or HM- sync points foreshadowed by the opening four measures.

Hasty's and Mirka's projection models enliven metric sync analysis and map well with Principle #2's requirement for a minimum of three equidistant sync points to elevate an R- to M-sync point. The constant orientation and reorientation of sync type trajectories create a dynamic connection between any two given sync points. The tripartite model further organizes audio streams into comprehensible groups of identical sync types. Understanding both the predictability and unpredictability of unfolding sync points deepens the relationship between organizational (hyper)metric and disruptive rhythmic sync functions. This projective analytical framework creates trajectories that I term "sync trains," to be detailed in the next chapter.

4.4: Focal Case Study

This chapter's focal analysis explores how visual accents may correlate to musical harmonies and accents in a way that frames larger audiovisual narratives. The case study of Sergei Prokofiev's "Battle on the Ice" cue for Eisenstein's *Alexander* Nevsky demonstrates how application of my sync point analysis technically describes a variety of rhythmic, metric, hypermetric, sync points interactions in real time.[123] Such interactions include how R[x]-[y]

[123] At the time of this analysis, I am unaware of any English-language analysis of the "Battle on the Ice" cue that walks through the entire cue by studying synchronicities. The cue was originally featured and self-analyzed by Eisenstein in *The Film Sense*, according to his concept of *audiovisual counterpoint*, the total montage-based interaction of sound and image. My analysis analyzes the same passage according to my sync point analysis framework, with respect to Eisenstein's original analysis. Differences are based on the definitions and limitations set for my terminologies. Most analyses, including Kevin Bartig's, stops after the first eight measures. See Kevin Bartig, *Sergei Prokofiev's Alexander Nevsky* (Oxford: Oxford University Press, 2017).

Even in Roy Prendergast's composition-oriented guide to film music glosses over the particulars of the scene, dismissing the validity of the graph because of Eisenstein's conflation of synchronous points. Prendergast criticizes the audiovisual score as ineffective because Eisenstein was too concerned with the content of the music and measures as opposed to the relationships of "picture rhythm." In other words, the graph became a conflation of vertical and horizontal image-sound relationships. While I agree with Prendergast to some extent, I argue that the audiovisual score is most useful when its production properties are transformed into analytical criterion for

building blocks both advance and disrupt the metric M[x]-[y] flow of rhythmic audiovisual editing. HM[x]-[y] sync points emerge at certain points in analysis to clarify the audiovisual narrative and nuance the three levels of sync point functions. The result is a hermeneutic sync point analysis that calls attention to accents of structural harmonies, chromaticism, rhythmic constancy, and metric disruption that create a total narrative of emergent meanings.

It is crucial to note that there are different versions of *Alexander Nevsky* available on the internet at the time of this writing, due to many re-recordings, remastering, and resynchronizations of the film.[124] Figure 4.12 outlines the sync points of the recordings that most closely resemble Eisenstein's audiovisual graph in *The Film Sense*, where the beginning of the music squarely matches with the beginning of the scene.[125] While a remastered version made

---

measuring sync points according to film rhythm. Pinpointed shortcomings of the original audiovisual score can reveal analytical philosophies and priorities that inform audiovisual analyses. However, I believe that transformational analysis can help clarify these conflations. For more, see Roy Prendergast, *Film music: A Neglected Art: A Critical Study of Music in Films* (New York: W.W. Norton, 1992).

For the original analysis, see Sergei Eisenstein, *The Film Sense*, trans. Jay Leyda (New York: Harcourt, Inc., 1943), 174–216. In Eisenstein's "Battle on the Ice" cue from *Nevsky*, he prioritized music as the audio component, contrasting how mainstream live action films of that time were primarily vococentric. This excerpt is generally referred to as the iconic "Battle on the Ice" scene, but Eisenstein's writings often refer to specific parts of the scene, such as "attack of the knights," "punishment of the knights," and "dawn of anxious awakening." When discussing the "dawn" or daybreak portion of the "Battle on the Ice," Eisenstein remarks the dawn precedes the "Battle of the Ice" proper. From the reading, it is unclear if Eisenstein was differentiating "dawn" from "Battle on Ice," or if the title represents a battle within the same cue that shares the same name. The "dawn" section represents the twelve visual shots and 17 measure of music and is considered part of the entire "Battle on the Ice" sequence. For the purposes of this dissertation, the "dawn" section will be referred to as the "Battle on the Ice" cue to acknowledge its role in the entire cue.

[124] In addition to the original 1938 release, *Alexander Nevsky* has two more re-releases. According to Kevin Bartig, a Mosfilm restored version was released in 1986 in a box set with *Ivan the Terrible*, parts I and II. For more, see: Kevin Bartig, *Sergei Prokofiev's Alexander Nevsky* (New York: Oxford University Press, 2017): 139. It is assumed that the Mosfilm restored version made available on YouTube in the year 2022 and referenced for this analysis is the 1986 version, despite the video description attributing the 1938 version. The video is available at the following citation: Mosfilm, "Alexander Nevsky | DRAMA | FULL MOVIE | by Sergei Eisenstein, YouTube video, 55:26, https://youtu.be/Gq4PaJfod4w?si=oUU2B9FokDotRiW6.

And according to IMDB, a 1995 version featuring re-recorded music of Prokofiev's score to reproduce the film in a way that it was "perfectly synchronized to the original 1938 dialogue and sound effects tracks, so that it is now possible to see and hear the film exactly as it always was, with the exception being that the music is now heard in hi-fi sound, rather than the tinny 1938 recording." For the IMDB description, see: The Internet Movie Database, "Alexander Nevsky: Alternate Versions," The Internet Movie Database, accessed June 2024, https://www.imdb.com/title/tt0029850/alternateversions/?tab=cz&ref_=ext_shr_lnk.

[125] What is believed to be the 1938 version is scarce and difficult to find and verify. There are two available videos of the version that are excerpted from the complete film, which is understood to have suffered deterioration over time. The version most closely cut to Eisenstein's audiovisual score can be found on YouTube at the following: Fix

widely available by Mosfilm is available, the rate of shot changes differ greatly from the audiovisual sync featured in Eisenstein's *Film Sense*.

The annotations in Figure 4.12 outline M[S]-[ISO] and R[S]-[ISO] as the two predominant sync point types.[126] These two types create an organizational "form" whereby the mid-section R[S]-[ISO] points (mm. 8–13) are bookended by M[S]-[ISO] collections. The R[S]-[ISO] that enters in m. 8 disrupts the metric constant established in prior measures, creating a reorientation of sync types and dynamizing the mid-section. The R[S]-[ISO] denies further projection of the preceding M[S]-[ISO] and is emphasized by the red arrow in Figure 4.12.

It could be said that the entire passage is made of R[S]-[ISO] sync types. However, the sync points at the beginning of Figure 4.12 form a referential metric constant in the opening bars of the cue. A black and white title card onscreen sets the scene to "April 5, 1242." What is initially an R[S]-[ISO] shot change recurs every two measures from mm. 1–8 and cuts to shots of the sky, two soldiers, and a flag. Though the M[S]-[ISO] occur every two measures and meet the minimum for am HM- sync, the sparse musical "phrasing" of the passage encompasses four measures. Therefore, the label of M[S]-[ISO] is preserved at the initial level to indicate durational equidistance without regard to phrase structure, saving the HM- label for four-measure units. These opening bars encourage a projection trajectory for the recurrence of M[S]-[ISO]

Me a Scene, "Alexander Nevsky (1938) — Battle On The Ice," YouTube video, https://youtu.be/IcPixaWL2Pg?si=jhsI-eePg7_hj8YS.

Alternatively, the "Battle on the Ice" scene can also be found uploaded to *Internet Archives*, with a music soundtrack that more closely resembles Prokofiev's later orchestration for his cantata version. See also, Russia, "Russia RU - 🎥 #OTD in 1938 Sergey Eisenstein's historical epic "Alexander Nevsky" premiered - to astounding success. It was a visual feast of unprecedented scale (battles!) with an OST by Sergey Prokofiev himself! 💬 Quote: "Whoever will come to us with a sword, from a sword will perish," Internet Archive video, 0:25, https://archive.org/details/twitter-1466145850166587401. The shot changes in this version also closely mirror the lines and sketches Eisenstein made in his *The Film Sense* graph.

[126] The reduction and transcription of all musical examples of "Battle on the Ice" moving forward references two main sources: the sparse music outline provided in Eisenstein's "audiovisual correspondences" in *The Film Sense*, and Prokofiev's cantata version. For the cantata version, see: Sergey Prokofiev, *Alexander Nevsky: Cantata for Mezzo-Soprano Solo, Mixed Chorus and Orchestra, Op. 78* (Tokyo: Zen-On Music Co., 1999).

sync marked by the blue arrows, and the two-bar M[S]-[ISO] maintains Principle #2's tripartite preference until the R[S]-[ISO] enters on m. 8.

The role of HM- sync is more nuanced in the "Battle on the Ice Cue" and coincides with melodic and harmonic aspects of the sparse film score. Illustrated in Figure 4.13, the HM-sync spans four measures each, accenting sync points (labeled S.P. in the annotations) 1 and 3. Here, the [x] variable for [music accent] changes from M[S] to HM[P] to emphasize the "phrase" at a larger-scale level. While the two HM[P]-[ISO] spans are not melodic in a traditional music sense, the semitonal motion serves as minimal indications for hypermetric phrasing. The bracketed semitonal motion in Figure 4.13 call attention to what Michael Lane Bradshaw identifies as Prokofiev's simplistic style. This style involves "Chromatic or stepwise voice leading with triadic harmony," and what Kevin Bartig regards as the half-step motion important to the cue.[127] The half-step is not only found in the upper voices, but also in the semitonal root relationship between the Cm and C#m bookending the excerpt. The semitonal C-B in the top voice highlights the intervallic motion that also recurs in sub-themes following the "Battle on the Ice" cue. Additional semitonal motion in the rising bassline of the cue add to the semitonal saturation in the passage and reflects Bradshaw's observation that Prokofiev's simplistic music also features "a preference for organizing both local and large-scale key relations around stepwise motion."[128]

---

[127] Michael Bradshaw, "Prokofiev's New Simplicity and Expressive Form," Phd. dissertation, (University of Queensland, 2018): 193–94. Bradshaw, among many other scholars, notes that Prokofiev's mid-1920's music shifted towards a sense of "New Simplicity."

Bradshaw remarks that new simplicity did not erase but rather minimalized evidence of Prokofiev's style. This includes, but is not limited to, use of polytonality, certain sets (4-20), hexatonic poles, chromatic and stepwise transformational voice leading, and "large-scale key relations around step-wise motion." On page 3 of his introduction, Bradshaw also proposes three reasons for Prokofiev's stylistic shift: (1) Prokofiev's return to Russia, (2) his "lack of success with some works such as the Second Symphony," and (3) his "joining of the church of Christian science around 1922-24…"

See also Kevin Bartig, *Sergei Prokofiev's Alexander Nevsky* (Oxford: Oxford University Press, 2017). Bartig identifies the "half-step motive" as evident in the cue.

[128] Bradshaw, 194.

At the surface level, Figure 4.13 does not have a tripartite construction but still follows Principle #4, where HM- type sync is not required to meet the minimum tripartite requirement if it is elevated from an existed metric sync. The pre-existing M[S]-[ISO] fulfills this condition as the HM[P]-[ISO] accents have only two occurrences and do not fulfill Principle #2. The first HM[P]-[ISO] begins with C in m. 2, moves downwards to the B in m. 3 (subsuming the M[S]-[ISO] that marks it), and is punctuated by a third leap down to a G# that is prolonged through m. 4. This semitonal motion is marked by the M[S]-[ISO] in m. 3. The same musical material recurs in mm. 5–8, reinforcing the phrasal unit presented in its first presentation in mm. 1–4.

Beyond the surface level, I argue that there are three inexact presentations of HM[P]-[ISO], albeit disrupted by a passage of rhythmic sync points. This aspect is realized through the harmonic dimension of the music. In addition to intervallic motion and phrase structure, the HM[P]-[ISO] sync also deepens sync point analysis by calling attention to Neo-Riemannian transformations underscoring the sync point changes.[129] Prior to the entry of R-sync (mm. 8–12), the annotated triadic transformations coincide with the HM sync in mm. 1–8. In measure 8, B-major to G# minor transformation leads to an increase of visual accents and stalling of music on the G# minor triad (mm. 7–11). As previously noted, complete sync between transformation, animation, and montage returns in m. 14 to balance and bookend the excerpt with the same HM-sync from mm. 1–8. The large-scale L' motion from the opening Cm to the GM following the G#m stagnation finally re-establishes the (hyper)metric sync points. This large-scale motion

[129] The annotated score is transcribed by the author, with reference to Prokofiev's Op. 78 cantata adaptation. The use of NROs to analyze Prokofiev's music has been previously explored by Olga Sologub, and the use of NROs in film music is established by Frank Lehman. My methodology combines both practices to understand the harmonic underpinnings of marked synchronous points.

For more, see Olga Sologub, "Sergei Prokofiev's Piano Sonata No. 8, op. 84 and Symphony No. 5, op. 100: Neo-Riemannian and Kholopovian Perspectives," PhD dissertation, (University of Manchester, 2014).

stresses m. 14's G#m, as if to correct the sync point and harmonic disruption through reinforcing downbeat compound PL operations.

The ending of the excerpt (mm. 14–17) manifests (hyper)metric sync's *superstructure* status and organizational function at sync point #9. Prior to this span, metric (mm. 1–7) and rhythmic (mm. 8–13) sync points are clearly presented one after another. In mm. 9–17, R-sync points 10 (m. 14) and 12 (m. 16) disrupt the metric potential of sync points 9 and 11. However, the final stretch of mm. 14–18 combines the metric accentuation from mm. 1–8 with the offbeat rhythmic sync in the rising theme. The musical material in the top line of the transcription recalls the opening phrase, connecting the possibilities of M[S]-[ISO] and HM[P]-[ISO] returning in the same manner as the first eight bars. The bottom features R[S]-[ISO] like that of mm. 8–13. The union of these two lines complicates analysis, but the hierarchical nature of the hypermetric sync allows the HM[P]-[ISO] to subsume the lower-level M[S]-[ISO] sync points #9 and #11 as well as the R[S]-[ISO] sync points #10 and #12.

To create a narrative summary: an audiovisual narrative arises where the first two iterations of HM[P]-[ISO] establish a projective instance ("tone") 1 and 2. Upon the third (mm. 9–13), rhythmic sync points deny the further projection of HM[P]-[ISO] and features R-sync over a stalled G# minor harmony. Sync points #5–8 (mm. 8–12) disrupt the HM[P]-[ISO] and accentuate a transposed version of mm. 1–8. This reiteration of familiar music is complicated by sync point patterns different than the initial one in mm. 1–8. The four R[S]-[ISO] points (mm. 8–12) are inconsistent and accelerate visual accentuation over a stagnant G# minor triad. Sync point #4 correlates with a metric downbeat and creates a false sense of restoration. However, R[S]-[ISO] one measure later and absence of M[S]-[ISO] in m. 9 denies a non-tripartite unit according to Principle #2 and retains the local level R-sync label.

Upon the fourth iteration (mm. 14–17), the HM[P]-[ISO] is reformed over the simultaneous return of R[S]-[ISO] sync points. Given HM[P]-[ISO]'s hierarchical and organization function, it is given the priority in analysis here, relegating other lower-level sync types M- and R- as subordinates. The framing HM[P]-[ISO], is however, emergent from what would have been R-sync types and does not have three exact repetitions. Yet, in accordance with Principle #4, HM[P]-[ISO] is achieved through elevation from their M-sync building blocks. The resulting analysis places importance on m. 14–17 as bookends that feature an HM-sync passage to "correct" the missing "tone 3" expected four measures earlier.

There is no clear-cut [AVC] cadence in this example. The transformative nature of NRO's allows the absence of traditional V-I cadential activity. This [AVC] absence heightens the importance of the metric sync types bookending the excerpt and almost creates an imbalanced A-B-A' form for the cue. The emergent framing structure organizes the audiovisual interactions and encourages potential sync pattern projections. In the end, sync point analysis not only reveals the technical artistry of the "Battle on the Ice" cue, but also glues together otherwise non-recurring in-shot content. Therefore, (hyper)metric sync plays a large role in defining R-, M-, and HM-relationships as well as organizing larger musical passages and structuring an audiovisual narrative that backbones the filmic experience.

4.5: Conclusion

Chapters 2–4 establish a full analytical and taxonomical toolkit. The sync point framework can now analytically illustrate the active filmic experience that unfolds in time while building cyclic references to past and present that inform the future. This energistic approach to

sync point analysis unfolds through the confirmation or denial of projected metric sync points in tandem with the inconsistencies of disruptive rhythmic sync points.

Audiovisual sync analysis can help understand what Rick Altman notes as "a story of that event," as opposed to only focusing on the events of that story.[130] Within the analytical framework, (hyper)metric sync points carry a reciprocal and simultaneous relationship with R-type sync points. Rhythmic sync points continuously build potential metric relations that are either fulfilled or denied through tripartite realization. Elevation to metric sync types calls attention to organizational accents, while maintaining rhythmic sync status highlights dynamic accents that push action forward. With two streams of sync point possibilities always evident, audiovisual experiences weave in and out of different sync types to create narratives through technicality.

Through rhythmic sync, internal rhythms tend to correlate with moving surface rhythms at the horizontal plane, filling in moments of stable shot changes to avoid any sense of visual stagnation during harmonic prolongations. As with the G# minor harmonic stall in "Battle on the Ice" cue, harmonic prolongations serve as stagnant points where active between- (shot changes) or in-shot (character and camera motion) film rhythms avoid audiovisual stagnation. This impetus may lead to asynchronous interactions on the local level that are synchronous on a hypermetric level. This is the case for the metric sync points in mm. 9–17 of the "Battle on the Ice" cue that subsume the two nonperiodic rhythmic sync points in the same span. In the next chapter, the sync types and five principles offer a peephole into the intricate interactions of audiovisual synchronicity theorized by Eisenstein, perpetuated in modern films, and analyzed through my sync point analysis framework.

---

[130] Rick Altman, "General Introduction: Cinema as Event," in *Sound Theory, Sound Practice*, ed. Rick Altman (New York: Routledge, 1992).

The next step in sync point analysis is to take the full toolkit of R-, M-, and HM- sync points and apply them to projection graphs. This process creates streams of metric projection I term *Sync Trains* that summarize the total function of an audiovisual scene in a digestible visual graph. The graphs will, in turn, pinpoint the audiovisual practices that push the filmic experience beyond its purely phenomenological notions.

Part II: Advanced Applications

Chapter 5

Dynamic Sync Trains

5.1. Sync Trains

The sync point analysis framework built in Chapters 2–4 helps define two aspects of audiovisual relations: (1) differentiating synchronous from nonsynchronous interactions and (2) labelling what type of synchronous functions are at play.[131] To expand the framework's capabilities, this chapter introduces the concept of sync trains to describe what happens *between* sync points. The sync trains articulate technical methods and functions to demonstrate how rhythmic profiles built from analyzing sync points intensify and energize narratives. Sync trains will additionally clarify two key aspects of said undefined spaces: (1) whether the in-betweens connect sync points with metric periodicity (regularity) or rhythmic non-periodicity (irregularity) and (2) *how* the in-betweens are audiovisually presented to viewers. The in-betweens are described by three terms that indicate equal use of audio and visual techniques (totality), more salient visual accentuations (visuality), or more salient musical accentuations (aurality). Combining sync point analysis with sync train construction results in a scoreless map that visually recreates the dynamic filmic process on paper. It also graphically illustrates how sync points can bridge small and large durational spaces to develop a rhythmic profile. This rhythmic profile is independent and not solely determined by the music nor visual cutting. Instead, it emerges from collections regular and irregular synchronicities arising from interacting musical and visual accents.

[131] This is defined by the information input into the n[x]-[y] formula. Please refer to Chapter 2 for the possibilities.

Figure 5.1 presents an abstract sync train made of atemporal nodes, lines (straight/wavy), and markers. Each visual component is defined as follows:

1. Nodes represent sync points.
    a. Filled nodes indicate use of shot change.
    b. Unfilled nodes indicate substitution using character/camera motion.
2. Lines represent "space" in between each sync point.
    a. Straight lines indicate regularity (metric projection).
    b. Wavy lines indicate non-metric irregularity (non-metric projection).[132]
3. An hourglass marker signals an upcoming *change* from one sync type to another.

Aside from these three building blocks of sync trains, there are accessory annotations such as Sp. for "sync point numbers" and time stamps/measure numbers to pinpoint the location of a synchronous event. Adding information to the train creates a segment like that of Figure 5.2's sync train and accompanying transcription of a passage from "Close to Gray," a music video from the *Hatsune Miku: Colorful Stage!* franchise.[133] Each Sp. 1–4 in the transcription is transformed into a node. The nodes are labelled R[D]-[ISO] according to the sync point labelling system and indicates how each sync point is non-periodically marked by general camera and

[132] Therefore, wavy lines indicate the continued projection of non-metric sync point types. While Chirstopher Hasty's original projection model deals solely with metric projection, the sync train allows even non-metric units—specifically [R] type sync—to be "projected" per the definitions of sync point analysis.

[133] The transcriptions of "Close to Gray" used in this dissertation is done by the author, by ear. However, an additional fan transcription score was referenced to double-check pitches. There are some differences between the author's transcription from the fan transcription, one being the indication of swing rhythm and another being transcribing only relevant musical parts and including lyrics. The fan transcription is arranged for solo piano performance and includes additional harmonies and embellishments for performance purposes. For the fan transcriptions, see: Surii (すりぃ), 限りなく灰色へ (Close to Gray), arranged by ShiA, in *Musescore.com*, accessed June 2024, https://musescore.com/user/34905762/scores/6656120.

The media reference for "Close to Gray" is as follows: HATSUNE MIKU: COLORFUL STAGE!, "HATSUNE MIKU: COLORFUL STAGE! - Close to gray by Three 3D Music Video - Nightcord at 25:00," YouTube video, https://youtu.be/3054q5rA5JQ?si=OgQQa3gLe1JeN7tS

character motion such as characters entering the frame (Sp. 2), turning their heads (Sp. 3), and initiating the dance choreography (Sp. 4).[134] These specific details are then generalized and converted into the unfilled white nodes of the sync train to reveal how the underlying idea of "motion" connects these four sync points. The wavy lines indicate that the *space* and audiovisual material connecting these nodes are non-periodic, per the R[D]-[ISO] label.

By defining functions of in-between spaces (represented by lines), sync trains provide foundational data that build context for otherwise disparate sync point functions. Sync trains can also involve more complex ideas regarding synchronicity and the construction of narrative through alternations or absences of periodic and non-periodic passages. This is possible because of the energistic nature of sync trains and its ability to graph fluctuating audiovisual rhythms. In doing so, the trains invoke the music theories of David Lewin, especially Lewin's *transformational attitude* that positions two harmonies as points A and B. By clarifying regular and irregular events in connective spaces between points A and B, sync trains can additionally invoke Hasty's, Mirka's, and London's music theories that explain listeners' tendencies to shape musical experiences through understanding of time, rhythm, and meter.[135] At the same time, sync trains can also keep track of and determine how visual accents form a rhythmic narrative that may influence viewer expectations.

In her chapter "Musical and Film Kinesis," Danijela Kulezic-Wilson also cites Eisenstein's theories on audiovisual synchronicity:

---

[134] R[D]-[ISO] is the sync type label that indicates the following: R = *rhythmic*, non-periodic synchronicity. [D] for [Duration] indicates that the "duration" of the music mirrors the duration of the visuals. The result is a mirroring of durational structure that is [ISO], or [Isomorphic]. This label follows the n[x]-[y] formula as outlined in Chapter 2, whereby n defines the *rate of occurrence,* and [x]-[y] respectively defines [musical accent] and [visual accent] types.

[135] While Mirka argues that meter is established by composers through the score, London provides a framework of cognitive templates and entrainment with the belief that meter is not always notated. For more, see Justin London, *Hearing in time: psychological aspects of musical meter* (New York: Oxford University Press, 2012).

See also, Danuta Mirka, "Musical Meter between Composition and Perception," In *Metric Manipulations in Haydn and Mozart* (New York: Oxford University Press, 2009), 3–30.

> "…identifying movement as the common denominator for music and film does not solve the problem of establishing a consistent methodology for the investigation of audiovisual movement – if that is possible at all – because synchronization between sonic and visual components constantly fluctuates between various audiovisual levels."[136]

To address this problem, my sync train framework seeks to provide a methodology that can track fluctuations of audio and visual markedness. It does so by organizing different sync points into collections and tracks their markedness via periodicity and the level of participation (contextualization) among the audiovisual components. In doing so, sync trains also reconstruct the inner workings of a scene or video to underscore what Karen Pearlman considers the editor's tools for crafting film rhythm: "time, energy, and movement" that push beyond notions that film is a purely phenomenological experience.[137] As a precedent for these possibilities, the following "methodology" section outlines the methodology for constructing sync trains that, at its core, can assess and define connective spaces between two sync points. Afterwards, the focal case study of this chapter, "Close to Gray" by Nightcord at 25:00 demonstrates how the trains contribute to building a more complete picture of the filmic experience.

---

[136] Danijela Kulezic-Wilson, *The Musicality of Narrative Film* (Basingstoke: Palgrave Macmillan, 2015), 84. See also: Eisenstein, *The Film Sense*, 168. Eisenstein's focus on connecting shots and "movement" is central to the entirety of *Film Sense*. Throughout his book, he writes passages such as the following: "From the practical material presented above we can formulate a simple, practical approach to a method of audio-visual combinations. *We must know how to grasp the movement of a given piece of music, locating its path (its line or form) as our foundation for the plastic composition that is to correspond to the music*."

[137] Karen Pearlman, "On Rhythm in Film Editing," In *The Palgrave Handbook of the Philosophy of Film and Motion Pictures*, eds. Carroll Noel, Laura T. Di Summa, and Shawn Loht (Cham: Springer International Publishing, 2019): 146–49, https://doi.org/10.1007/978-3-030-19601-1_7.

5.2: Sync Train Methodology

5.2.1: Background

Sync trains build upon the sync point analytical framework outlined in Chapters 2–4 by connecting the points through imagined space. Therefore, the analytical conceptualization and recreation of imaginary *motion* underlies the construction of sync trains. The principles behind the graphs are influenced by a combination of David Lewin's writings dynamic phenomenology and transformation theory from *GMIT* (1987) with Sergei Eisenstein's montage ideas from *The Film Sense* (1942).

In his 1987 publication, music theorist David Lewin used a line graph to describe dynamic intervallic thinking as pre-transformational concept.[138] Lewin explains the intervallic thinking as a dynamic perception of an imaginary "line *i*" that "symbolizes a characteristic directed measurement, distance, or motion from [points] s to t."[139] In Figure 5.3a, I map the line *i* from Lewin's preliminary graph onto film studies concepts to demonstrate abstract movement from one sync point to another. This intervallic movement mirrors the durational connection (line *i*) between image-contents of shots A and B. Lewin ultimately departs from intervallic thinking to develop his *transformational attitude*, where shot A *becomes* shot B instead of simply moving from shot A to shot B. While film studies itself does not have transformational functions in the manner of music theory for changing shot A into shot B, it is the very idea of a dynamic line *i* that can lead from point A to point B *or* transform point A to B that represents the focus of dynamic sync trains.

---

[138] David Lewin, *Generalized Musical Intervals and Transformations*, New Haven: Yale University Press 1987.
[139] Ibid., xxix.

To address the absence of a proper set of transformational functions for film studies, I invoke Lewin's dynamic phenomenological thinking that is independent but complementary of his *transformational attitude*. In his writings on Schubert's *Morgengruß*, Lewin provides a very technical approach to explain the perception of music, one that is determined by the context of otherwise independent musical notes or "acoustic objects" and then directed by a logical goal, such as a basic I-V-I harmonic motion.[140] The percepts, or musical objects, create worlds and ecosystems of musical functions and harmonic vocabularies within the confines of a composition, and the true functions of voice leading and harmonies unfold in real time while also reinforming past musical events.

In this vein, I take shots A and B in Figure 5.3a to represent two percepts, and the line *i* to indicate the *act* of audiovisual materials contextualizing sync point projections defined in Chapter 4. In other words, sync trains rely on the following combination to visualize and recreate the dynamic film experience on paper:

Perception + Projection = "transformational" filter

Here, perception stands for the perception of audiovisual data and contexts surrounding specific sync point pairs, while projection refers to informed metric projection via the dynamic theories of rhythm and meter discussed in Chapter 4. The transformational filter formed by the combination of perception and projection helps relate sync point A to sync point B by making

---

[140] David Lewin, "Music Theory, Phenomenology, and Modes of Perception," in *Music Perception: An Interdisciplinary Journal* 3, 4 (1986): 327–92. https://doi.org/10.2307/40285344. See page 353 for Lewin's discussion on the differences of percepts being "two distinct mental objects" or the contrasting "one acoustic object."

Lewin's dynamic phenomenology and his 1974 essay on *Morgengruß* is later formalized for publication with more depth and detail. See also: David Lewin, with David Bard-Schwarz and Richard Lawrence Cohn, "Morgengruß by David Lewin," in *David Lewin's Morgengruß: Text, Context, Commentary*, eds. David Bard-Schwarz and Richard Cohn (New York: Oxford University Press, 2015): doi: https://doi-org.uc.idm.oclc.org/10.1093/acprof:oso/9780199844784.003.0002.

line *i* a contextualization of framed audiovisual happenings that uses projection to advance and retrospectively reinform R-sync, M-sync, and [AVC] types. Therefore, sync points serve as percepts A and B that use "perception + projection" to transform the filmic experience. To this end, Chapter 5 focuses on periodicity (regular/irregular) and presentation of audiovisual contexts (totality, visuality, aurality) as key factors of perceiving and forming sync trains.

Figure 5.3 also invokes Pearlman's chapter on film rhythm as both a poietic and esthesic "*felt* phenomenon," that "Putting two shots together, each of which inherently has rhythm, makes a third rhythm, which is not the same, or even just the sum of the first two."[141] Sync trains focus on contextualizing the "third rhythm," or the emergent rhythm that arises when music and image meet and are connected by in-between materials represented by my adaptation of an imaginary line *i*. This plays into Brian Kane's observation that Lewin's dynamic phenomenology model is based on both perception and "performative, creative, and generative" action.[142] Therefore this dissertation posits that the active reception of sync points is the dynamic within perceived contexts of audiovisual data.

The harmonic transformations that Lewin's transformational theory addresses are atemporal. The points *s* and *t* in Figure 5.3a can be reinterpreted as harmonies A and B, whereby harmony A is changed to harmony B through a musical action (i.e. "transformational filter") that is line *i*. For example, a C major triad (harmony A) can transform or change to an A minor triad (harmony B) through a "relative" relation (line *i*) that moves the fifth scale degree of C major up

[141] Pearlman, 143 and 146.
[142] Brian Kane, "Excavating Lewin's 'Phenomenology," in *Music Theory Spectrum* 33, 1 (2011): 27, https://doi.org/10.1525/mts.2011.33.1.27. Kane critiques and summarizes Lewin's article "Music Theory, Phenomenology, and Modes of Perception," and links Lewin's ideas to that of West Coast phenomenology. Near the beginning of the article, Kane summarizes that "By drawing together perception and action, he discourages the reader from conceiving of perception as a disembodied process—it is not simply receptive, but rather performative, creative, and generative."

by two semitones, while preserving a shared major third (music notes E-G) between the two chords. These transformations happen abstractly, and not in real time (retrospect) with the music.

On paper, Lewin's focus on actions that take the listener from one point to another seemingly conflicts with the sync trains' focus on describing rhythmic motion and tracking emergent rhythms in real time. However, sync trains are constructions of atemporal possibilities in filmic space. The nodes presented in the trains serve as points A and B, and the straight and wavy lines mirror Lewin's dynamic line *i* by placeholding for one of three possible relations discussed later in this chapter: totality, aurality, and visuality.

Sync trains seek to describe how an audiovisual frame such as a stationary character becomes the next frame, like a dancing character, through an action. At the same time, the framework focuses on describing the motion, duration, and chronological progression of events. This is where Eisenstein's montage theories are incorporated to create a more encompassing framework that accounts for points A and B as well as the action that takes viewers from one image-point to another.

The concept of building connections is important to Eisenstein, although his early theories were more interested in the relations among static images A and B instead of the action or motion that takes viewers from A to B. In the chapter "Synchronization of the Senses" from his book *The Film Sense* (1942), Eisenstein notes that images A and B are connected by dialectic relationships.[143] Said dialectic relationships and emergent meanings are created by Eisenstein's montage techniques, whereby "*Representation* A and *representation* B must [through their *juxtaposition*] …. evoke in the perception and feelings of the spectator the most complete *image*

[143] Sergei Eisenstein, *The Film Sense*, translated by Jay Leda (New York: Harcourt, Inc., 1943): 69.

of the theme itself." [144] I apply Eisenstein's idea to the audiovisual nature of sync trains, where two distinct entities of image and sound can come together to deliver the theme, or a singular message that all filmic elements point towards. However, the rise of sound films poses a compositional problem in the interactions between A and B and leads Eisenstein to theorize synchronicity as the "*measured matching* of a strip of music and a strip of picture" as a viable solution in reconciling the otherwise separate audio and visual components.[145] The measured matching creates "vertical" relations among image and audio, while forming "horizontal" connections between images A and B.

The sync trains take Eisenstein's images A and B determined by a film/video and equate them to Lewin's points A and B determined by the music. The sync trains then push the relationships of images A and B further by visualizing the imaginary *motion* between the points. In other words, the trains fuse Lewin's and Eisenstein's ideas together for analytical purposes. Though Eisenstein did not explicitly write of a transformational attitude, his viewpoint of how

---

[144] Eisenstein, *The Film Sense*, 69. The full citation reads: "*Representation* A and *representation* B must be so selected from all the possible features within the theme that is being developed, must be so sought for, that their *juxtaposition*—the juxtaposition of *those very elements* and not of alternative ones—shall evoke in the perception and feelings of the spectator the most complete *image of the theme itself*.

Eisenstein's writings in *the Film Sense* and later writings in *Film Form* emphasize the creation, emergence, and realization of an "organic whole" created by relations and interactions between points A and B. In *Film Form*, Eisenstein publishes an analytical essay on Japanese Kabuki theater to present the idea of a "monistic ensemble," or the staging of a unified, collective narrative that is realized by the equal contribution of staged elements. These "elements of equal significance" are identified by Eisenstein as sound, movement, space, and voice—all components of Eisenstein's conception of audiovisual montage and the *movement* between the imagery and themes presented by Representations A and B.

[145] Eisenstein, *The Film Sense*, 157. Eisenstein writes that audiovisual possibilities accompanying the rise of sound film can be solved by synchronicity, or "*measured matching*." He furthers that "…such measured matching as would enable us to unite both strips "*vertically*" or *simultaneously*: matching each continuing musical phrase with *each phase* of the continuing parallel picture strips—our shots. This will be conditioned by our adherence to the letter of that law allowing us to *combine* "*horizontally*" or *continuously*: shot after shot in the silent film—phrase after phrase of a developing theme in music."

This paragraph is Eisenstein's long form of describing the audiovisual synchronicity covered in this dissertation. Most important in this citation is Eisenstein's stress on not only uniting the vertical audiovisual sync points, but also the horizontal and continuous aspect of synchronicity—one that adheres to the linear, unfolding nature of music-listening and film-watching. I believe that this emphasis on the *horizontal* pinpoints what I consider Eisenstein's "transformational attitude." It is the continuous matching of "phrases" and "phases" that connects the otherwise static Representations A and B.

montage *transforms* the filmic experience complements Lewin's ideas. If I refocus his theories away from iconography, ideograms, and shot content, Eisenstein's rhythmic editing (i.e., montage) can be studied for how shot A's content is transformed into shot B through rhythmic cutting and stitching, or through modern techniques such as character and camera motion. Revisiting Figure 5.3a, montage relations from shots A to B interact in a dynamic manner like Lewin's example of his line *i* connecting points *s* and *t* in musical space. In my analytical framework, the line *i* represents undefined spaces, or audiovisual *actions*, between sync points.

Both the sync train nodes and the spatial intervals between Lewin's points A and B are inherently atemporal. Only when filled with information (i.e., sync point labels) do they form a sync train out of limitless possibilities in Pudovkin's "filmic space."[146] Figure 5.3b abstractly illustrates sync point possibilities in my adaptation of said "filmic space." Once an undefined space, the connective line now represents the finite boundary of time specific to each music video or film. The disparate dots represent any of the possible sync types introduced in Chapters 2–4. Only when the dots are placed and connected on the finite lines do they become purposeful and functional. Therefore, sync points drawn from these possibilities are organized and connected to determine the content and function of each sync point, while the imaginary line *i* serves as the *undefined spaces* between each sync point possibility. Defined in Figure 5.4, finite and organized interactions of sync points and in-between spaces create hierarchies of ordered sync point groupings:

- Sync *points* are basic building blocks with functions defined by sync types.[147]
- Sync *collections* are groups of recurring sync points of the *same* sync types.

[146] Vsevolod Pudovkin, *Film Technique and Film Acting* (New York: Lear, 1949).
[147] The sync labels refer to the n[x]-[y] framework. The n[x]-[y] is a technical term for "sync type" and is defined as *rate of occurrence* • [musical accent type]-[visual accent type].

- Sync *trains* are larger groups of same *and different* sync collections.
- In special cases, sync collections may feature a secondary sync type and create momentary "reroutes."

As the largest structure in the hierarchy, sync trains quickly elucidate different audiovisual functions by defining all types of musical change that surround and include individual sync points. Said musical change includes harmonic/melodic accents, metric projection, or rhythmic irregularities when they are present, making the framework flexible in charting out the dynamic film process.

Figure 5.5 is a scoreless summary of the "Battle on the Ice" cue from *Alexander Nevsky* (1938). The three large circular nodes are graphed to represent the initial type of each respective sync collection M[S]-[ISO], R[S]-[ISO], and a recurring M[S]-[ISO]. The black nodes indicate that all sync points were traditional shot changes. Lines are added to connect the sync points across filmic space, representing the train "rails" that connect the "train" graph together.

5.2.2 Periodicity: Regular and Irregular

Determining the characteristic of the periodicity as "regular" or "irregular" constructs a rhythmic profile drawn from animated musical performance. The periodicities elucidate the rhythms that emerge from synchronous moments as well as the function that *time* and consistency play in connecting disparate sync points together. It also demonstrates how sync collections determine the stability and rhythmic intensity of the narrative. Shown in Figure 5.5 are three different line ("rail") types that indicate metric synchronicity: (straight lines), rhythmic synchronicity (wavy lines), and "reroutes" that feature that include more than one sync type

within a collection (wedging lines). In the grand scheme, the lines indicate two modes, regularity and irregularity, to elucidate sections that are dominated by the projection of metric sync types (regular) or the projection of non-metric types (irregular).

Sync points 1–4 feature a straight line to indicate the presence of "regularity": metric sync points. The line also marks the continued presence of metric projection, per the frameworks of Hasty and London discussed in Chapter 4. This means that each small, filled black dot that follows the initial large node is a confirmed and continued M[S]-[ISO] sync point. Metric projection is thwarted only at the hourglass marker that signifies the end of one collection and an anticipation of another. It is a symbolic representation of an unfulfilled projection M[S]-[ISO] point and reorientation of a new projection path.

The wavy lines connecting SP. 5–8 contrast straight lines to indicate "irregular" non-metric projection, or the anticipation of any rhythmic [R] sync type that is not (hyper)metric and unpredictable. In this passage, the initial R[S]-[ISO] is projected, confirmed, and continued until the hourglass marker preceding SP. 9. Again, all sync points following the initial large node are identical to the initial R[S]-[ISO] point. The unpredictability signified by wavy lines also detail the audiovisual rhythm's intensification through non-periodicity.

"Reroutes" are irregular wedging lines that indicate brief detours to sync types different from the main collection.[148] As shown in SP. 9–13 of Figure 5.5, there are two wedges at SP. 10 and 12 labelled R[S]-[ISO]. These moments indicate that there are sync points other than the M[S]-[ISO] of the sync collection. However, the overall sync collection does not change because the reroutes are brief moments that do not override the salient type indicated by the initial node.

[148] To associate with music-theoretical terms, a *reroute* functions similarly to harmonic tonicizations, whereby different key areas are briefly highlighted without the music changing (i.e. modulating) to an entirely different key. Like tonicizations, reroutes maintain the salient sync type ("music key") while featuring brief highlights of different sync types.

This is because the reroutes that are *rhythmic* [R]-type sync cannot override the hierarchy of the salient metric [M] sync type. Additionally, the [R]-sync's non-periodicity cannot achieve the tripartite preference stated in Chapter 4: that "Metric sync points must accent at minimum three equidistant occurrences." Therefore, the reroutes are subsumed into the M[S]-[[ISO] collection but still acknowledged by the wedging lines found in Figure 5.5. Despite wedging from a straight line that indicates metric sync, reroutes are irregular line types because they disrupt the singular projection of metric sync types. Though not exemplified in Figure 5.5, reroutes can also wedge from wavy lines.

The three types of lines instantly indicate timespan periodicities as irregular (wavy/wedging) and regular (straight). Figure 5.6 demonstrates how lines also directly correlate to the numerical value of timestamps and/or measure numbers accompanying a sync train graph. The abstract train in Figure 5.6 features both rhythmic [R] and metric [M] collection. The wavy lines that indicate non-periodicity align with the uneven numerical values such as 0:19 (SP. 2) and 0:26 (SP. 3). The wavy lines can even match multiple sync points over a single timestamp, just as how SP. 4–8 happens within one second (0:27). This is contrasted by the straight line for metric sync points 9–13, where all the timestamp values are equidistant and occur every two seconds. Therefore, the steady timestamps support the fact that straight lines adhere to Eisenstein's definition of metric being of consistent, measurable values.

5.2.3 Presentation: Totality, Visuality, and Aurality

Sync trains can account for any form of musical accent including, but not limited to, harmonic, melodic, metric, or rhythmic change. While lines defining regular/irregular periodicity determine a sync train's rhythmic profile, shaded blocks provide the *contextualization* of sync

points. The blocks summarize the audiovisual content of the lines and contextualize sync points by assigning functions to the in-between spaces. This is determined by the levels of involvement that music and image play in presenting onscreen information according to a sync train's rhythmic profile. As shown in Figure 5.7, color shaded blocks encompass stretches of audiovisual action that include both sync points and their in-betweens (periodic lines). The following section outlines three general contextualization schemes that describe how the in-betweens—initially "undefined" spaces—function:

1. *Totality*: Contextual spaces that feature equal *non-coincidental* musical [x] and visual [y] markedness.
2. *Visuality*: Contextual spaces that feature more *non-coincidental* visual [y] than musical markedness.
3. *Aurality*: Contextual spaces that feature more *non-coincidental* musical [x] than visual markedness.

The emphasis on "non-coincidental" foregrounds the cyclic relationship between sync points (coincidental) and their contextual spaces (non-coincidental). While passages include both sync points and their contexts, the latter encompasses simultaneous musical and visual aspects that do not need to fulfill all definitions of synchronicity laid out in Chapters 2–4. In other words, there may be coincidental moments in contextual spaces, but such coincidences are not marked or accented enough to be a sync point.

Coincidental factors do have their own hierarchy and overall presence of markedness, which determines the salient functional element (audio, visual, or both) of the contextual space.[149] As

[149] These relationships are like [ICO], [ISO], and [I^2] but cannot be assigned these functions since contexts are not sync points.

this chapter's examples will show, it may be the case that a contextual passage features more visual action such as constant choreographed dancing between sync points (visuality), especially over a stretch of sparse music. In other cases, musical action such as rapid melodies or can be more emphatic than a sparse amount of visual action (aurality). Totality marks moments of oversaturated musical and visual action that are simultaneous, yet non-synchronous. For example, a passage with constant camera and character motion may happen at the same time as marked musical action *without* creating sync points on every possible beat. The action is better understood as a whole: that all elements are horizontally moving from one sync point to another instead of forming an oversaturation of verticality.

Totality's saturation of audiovisual data may overwhelm the senses and blur clear indications of musical or visual accents. Therefore, totality is best understood as a context where all parts, accented and unaccented, work together to deliver a singular emergent idea. For sync trains, the emergent idea manifests in a rhythmic profile formed from audiovisual interactions, a relation separate from the individual rhythms of music or visual editing.

In defining its function, my concept of totality adapts Sergei Eisenstein's idea of the "*monistic ensemble*," whereby all audiovisual elements create a unified, collective narrative (i.e., rhythmic profile) that is realized by the equal contribution of staged elements sound, movement, space, and voice.[150] In this chapter, I use sync trains to visualize and graph the monistic ensemble

---

[150] Sergei Eisenstein, *Film Form: Essays in Film Theory*, trans. Jay Leyda (New York: Harcourt, Inc., 1977), 20. *Film Form* includes Eisenstein's analytical essay on Japanese Kabuki theater and details his idea of a "monistic ensemble" and its staged elements of sound, movement, space, and voice. In this chapter, Eisenstein positions theater as the origin of film and cinema and reflects on theater via a dialectic between his knowledge of Western theater and that of the Kabuki performances that toured Russia sometime in 1928 (page 18). Eisenstein also argues against simply viewing Kabuki's dance drama as a foreign novelty and encouraged a study of the art form's provocative "*monism of ensemble*" that results in "a single unit of theater" (page 21).

In this chapter, I use sync trains to visualize and graph the monistic ensemble that creates emergent audiovisual rhythms. Sound and movement, as well as the hierarchical voice and [VCO] relationships, are determined by sync point definitions outlined in Chapters 2–4. The "space" is the contexts covered in the chapter. That being said, the contexts themselves possess sound, movement, space, and voice. It is the markedness of audio

that creates emergent audiovisual rhythms. Sound and movement, as well as the hierarchical voice and [VCO] relationships, are inherently determined by sync points. The “space” is the context types covered in the chapter. The contexts themselves possess sound, movement, space, and voice. But it is the markedness of audio (sound and voice) and visual (movement) that determines the type of in-between contexts used to spatially (space) connect sync points. Coincidence and the lack of synchronicity among the audiovisual parts prevent the formation of sync points while assigning totality to the in-betweens. Since all elements are active in totality contexts, and this overactivity results in little to no markedness of individual elements between sync points.

Combined with the periodic line markers (straight for metric and wavy for rhythmic), the three terms totality, visuality, and aurality turn “undefined” spaces into definable functions. These terms generalize how onscreen content functions in delivering the sync train’s rhythmic profile. This is especially important since the terms help describe the context and content surrounding disparate sync points. Stretches of totality, visuality, and aurality are not exclusive to any single sync collection within a sync train. In other words, one sync collection may feature fluid changes from one type of contextualization to another. Yet, it is the rate of occurrence and frequency that determines the dominant type of contextualization.

Color-shaded blocks are added to sync trains to easily differentiate the contextualization types. Moving forward, each of the contextualization types will be assigned a color as follows:

1. Totality = green shade
2. Visuality = yellow shade
3. Aurality = blue shade

(sound and voice) and visual (movement) that determines the type of in-between contexts used to spatially (space) connect sync points.

Very broadly, each category has definitive characteristics that will be expanded within this chapter. They also reinforce the sync point labels that they connect and are tied into the same assessment criteria that determines the R[x]-[y] and M[x]-[y] qualities outlined in Chapters 2–4. Totality is the most eventful of the three contextualization actions. It describes actions between sync points associated with constant harmonic changes, accelerated musical [x] accents, and accelerated rates of visual [y] accents that can either be R[x]-[y] or M[x]-[y] types. Visuality is most often defined by actions between sync points that feature static or slow harmonic changes, little to no musical [x] accents, and accelerated visual [y] accents. Aurality is the opposite of visuality, featuring accelerated harmonic changes, accelerated musical [x] accents, and little to no visual [y] accents.

Figure 5.7a revisits the "Battle on the Ice" cue from *Alexander Nevsky* (1938) to exemplify sections of totality and visuality. Distinct from the previous chapter's analysis, Figure 5.7a presents a complete sync train with all the exact information plugged in. Harmonies added below the measure numbers indicate the triadic transformations distinct to the excerpt. There are two color blocks, green and blue, that respectively represent totality and visuality. Aurality will be discussed in the following case study.

Green highlights marking *totality* encapsulate the A and A' sections and define how the in-between spaces feature audiovisual events where neither music nor image are more marked than the other. The in-between spaces better accommodate the imagined *motion* of harmonic transformations, or the change from one harmony to another that is not instantaneous but occurring *over time*. For example, the sync points at Sp.2, 3, 4, 8, 9, and 10 are all accompanied by a harmonic transform, be it an S, R, or PL. The sync collection spanning Sp.1–4 visually describes that visual accentuation by shot changes (indicated by black dots) synchronize with

harmonic transformations (harmonic annotations), held together by metric periodicity (straight lines and the M[x]-[y] label). The in-between events such as the C-B voicing in the topmost voice of the transcription connects the voice leading from shot changes Sp.1 to Sp.2, as well as the BM to G#m harmonic transform. The rising bassline also "moves" the music from the stagnant Cm harmony to the BM harmony (Sp.2) by means of an S (slide) transformation. Green shades in Figure 5.7 are added to indicate *totality* and areas where the in-between effects are equal parts music-harmonic and visual editing.

As a contextualization, totality also unifies a section that might have wedging lines. Seen in the A' section of Figure 5.7, the reroutes to R[S]-[ISO] does not impact the fact that each sync point correlates to a musical change such as harmonic transformation. For that matter, the inclusion of the two R[S]-[ISO] sync points (Sp. 10 and 12) respectively marking the initiations of the rising bassline and repeated G# drone strengthens the ascription of totality. It tells that the equal emphasis of musical and visual accentuation is present at the [M] and [R] hierarchies of the sync train's rhythmic profile.

The yellow color shade highlighting the B section (Sp.5–8) assigns *visuality* as the main type of contextualization. This is because the span features more visual than musical markedness. Due to the nature of sync points, there are still R[S]-[ISO] synchronicities that correlate to sonic aspects of the music such as the notes in the lower voices. However, *visuality* differs from totality regarding the number of musical parameters involved. In this case, all visual accents occur over a single musical accent: the G#m harmony. As the singular G#m harmony persists, it maintains a hierarchical markedness over the musical events that the R[S]-[ISO] accent. Therefore, the rate of shot changes seems relatively faster than the rate of musical change, even

though sync points inherently arise from synchronous events. Therefore, the term *visuality* is assigned to describe "contextual spaces that feature more visual [y] than musical markedness."

Aurality contrasts visuality by featuring more musical accentuation than visual markedness. Figure 5.8 illustrates the concept of *aurality* in an excerpt of "Close to Gray" (限りなく灰色へ) from the *Colorful Stage!* multimedia series. In Figure 5.8, the opening passage covers twenty-seven seconds of camera motion and sparse character motion. Compared to shot changes, camera and character motion are most emphatic at their precise point of initiation. For example, Sp.1 features a camera zooming out from a water fountain. The zooming resumes for seven seconds, but the motion—already set at Sp.1—is less emphatic and more so understood as a *continuation* of the initial Sp.1. The same idea applies to Sp.2, where the camera zooms out far enough for characters to enter the screen. Therefore, the annotations in Figure 8 call attention to subtle hand movements (Sp.2) and head turns (Sp.3) that may go unnoticed because they seamlessly blend into the general choreography. On the other hand, the harmonic and melodic actions are more marked. The repetitious melodic line unfolds in tandem with a recurring i^6-VI^7-III-VII progression, making the musical material more immediate and marked than the visuals. This is especially noticeable at Sp.3 and 4, because the characters move while the camera zoom, initiated at Sp. 1, is still happening. Therefore, the term *aurality* helps contextualize consistent melodic action and harmonic accentuation as the more marked in-between events that connect the disparate R[D]-[ISO] sync points.

Figure 5.8 brings about another key aspect of sync point analysis and sync train construction: that the rate of harmonic accentuation may differ from that of a harmonic progression defined by music theory standards. The looping i^6-VI^7-III-VII in this excerpt constitutes an "axis progression," Mark Richard's term for a looping progression "-a, -F, -C, and

-G" and its transpositions.[151] As the Roman numeral equivalent of -a, -F, -C, and -G, the i-VI-III-VII progression is harmonically static and tonic-prolongational. This means that, by music theory standards, the progression is as static and uneventful as the ongoing camera motion (visual action). The VI, III, and VI harmonies would not be functionally independent as they are supporting the long-term hearing of the tonic i chord. This would contradict the *aurality* label applied to Figure 5.8. However, since the sync point framework counts the rates of accentuation, the emergent audiovisual rhythm does not have to functionally match the music-theoretical harmonic progression (even though a mirroring of rhythms can be incidental to the framework). Therefore, the functional harmonic rhythm is different from the rates of harmonic *accentuation*, which is what each chord of the axis progression accounts for. In other words, each harmony of the looping i^6-VI^7-III-VII adds one accented sonic "hit" between sync points, even if the axis progression is functionally tonic-prolongating.

Looking at Figure 5.8, sync points occur only on harmonies III (Sp.1), i6 (Sp.2), and III (Sp.3), highlighting the axis progression's ability to simultaneously emphasize two possible tonalities i and III. The sync points themselves do not consistently emphasize a singular rotation of the axis progression, and nature of the R[D]-[ISO] sync means that the number of harmonic accents between sync points are unequal and nonperiodic. In the end, *aurality* is applied to this excerpt because of the active number of accents provided by both the rhythmic data drawn from the melodic and harmonic profiles of Figure 5.8. This tandem count of accents in between outweighs the number of visual accents occurring in this scene.

The totality contextualization schemes can also emphasize audiovisual cadences. Illustrated in Figure 5.9, *totality* highlights the ending of "Under the Sea" from *The Little*

[151] Mark Richards, "Tonal Ambiguity in Popular Music's Axis Progressions," in *Music Theory Online* 23, 3 (2017), doi: 10.30535/mto.23.3.6.

Mermaid (1989). Here, the high frequency of sync points demonstrates equal emphasis of musical and visual shot changes. Especially towards the beginning of the excerpt, the sync points are connected by constant musical and camera/character motion. The "total" acceleration of all parts creates an emergent rhythm to bring closure to a song. The in-between gaps disappear when the sync points increase in frequency. However, since sync trains include both sync points and their contexts, totality is applicable because all elements are working towards the singular function of cadential closure. This example exemplifies totality as the strongest scheme to emphasize formal closures.

Now that both periodicity and presentation (contextualization) are definable, sync trains will be tested through rigorous analysis in the next section. Sync trains' flexibility to accommodate any type of audiovisual change allows larger-scale analyses that include sync points *and* their surrounding context. This is facilitated by lines and shaded blocks assign function to once undefined spaces, building an emergent rhythmic profile out of sync points.

5.3: Focal Case Study: "Close to Gray"

Presented in Figure 5.10, the following "Close to Gray" case study will demonstrate sync trains' utility for analyzing a complete, choreographed music video. Beyond that, sync trains can also compare different adaptations of the same song. Comparing the two versions via sync trains is but a technical method of articulating how different adaptations of the same song intensify and energize narratives through rhythm. This is reliant on sync trains' ability to reverse-engineer the dynamic nature of a film or music video while highlighting the most important audiovisual elements that make a scene "work."

Analysis presents a set of problems that require sync trains to navigate fluctuations of periodicities and audiovisual markedness (contextualization). In addition to identifying sync points per the established framework in Chapters 2–4, problems such as competing sync types, multiple reroutes, and fluctuating markedness schemes are all managed and organized into a singular train. The ultimate purpose of a sync train is delivering a digestible graph that defines all audiovisual functions found in an animated music video.

Constructing sync trains is a form of reverse-storyboarding, or the reconstruction of focal audiovisual components that make a scene. Beyond providing data for individual scenes, sync trains provide graphs for comparison of poietic and stylistic choices among different film styles. This chapter's case study compares a commissioned Japanese popular song, "Close to Gray" (限りなく灰色へ), adapted across two distinct animated versions: one choreographed dance video with music sung by a five-member group and one traditional storyline narrative with music sung by a Vocaloid.[152] The original song is written by Vocaloid producer Three (also known as Surii or すりぃ) for the *Hatsune Miku: Colorful Stage!* series, and there are two official versions: the Nightcord version sung by live voice actors and Three's official version produced with the voice bank of virtual singer Kagamine Len. Despite differing voices and animations, both "Close to Gray" versions share the same dramatic and melancholic attitude.

Contrary to the upbeat backtrack, the melancholic choreography version emphasizes the lyrics that reflect Ena, a main character from the *Hatsune Miku: Colorful Stage!* series.[153] One of

[152] The term "Vocaloid" refers to Japanese popular songs that feature synthesized voices.
[153] Released in Japan as *Project Sekai: Colorful Stage!* (2020) and internationally as *Hatsune Miku: Colorful Stage!* (2021), this mobile app game is developed by various companies Colorful Palette, Craft Egg, and Crypton Future Media and is published by Sega. This dissertation will refer to all forms of the *Project Sekai: Colorful Stage!* games by its localized title, *Hatsune Miku: Colorful Stage!*, or simply by *Colorful Stage!*.

her personal arcs involves how Ena is unable to attend art school and always stands in the shadow of her unimpressed father, a famous artist in *Colorful Stage's* fictionalized Tokyo. With fruitless determination and hard work, Ena's worldview becomes miserable and grayed when her "talent is something borrowed," hence the title "Close to Gray." Her worldview is eventually "colored" with the help of her friends, and this transition from gray to color animation is reflected in the choreography version that represents her closure through song.

Figure 5.10 presents a complete sync train for the choreographed "Close to Gray" performance by the Japanese mixed media idol group, Nightcord at 25:00. All necessary information is immediately provided by the graph, but the following discourse will detail how each element contributed to the construction of the sync train.

There are four main sync types salient in the animated performance: R[D]-[ISO], R[S]-[VCO], R[S]-[I^2], and M[P]-[I^2]. Black and white nodes instantly indicate the use of onscreen motion (white) or shot changes (black) as the visual accent. The prevalence of wavy lines throughout the train indicates that Nightcord's choreography version is intensified by the oversaturation of rhythmic [R] sync types. This is because any type of [R] sync creates rhythmic tension due to its inherent non-periodicity and unpredictability. Likewise, the prevalence of black dots tells that, despite being a choreography video, the visual accents primarily use shot changes as a visual accent. This fact is particularly important since shot changes are more marked than character and camera motion that relies on the *initiation* sync point. The combination of frequent [R] sync and shot changes indicate that the choreography version is dynamic and event oriented.

The dynamic presentation of synchronicity is supported by the prevalence of the [VCO] type sync, especially for the large part of the music video (Sp. 5–35). Even without reference to the video, the sync train indicates that the frequent shot changes in the span of Sp.5–22 are

driven by a sense of vococentrism and, by extension, a commitment to sonic fidelity. Even when the sync collection switches to the R[S]-[I^2] collection at Sp.23, there are four reroutes to R[S]-[VCO] that maintain the video's commitment to sonic fidelity, a feature central to the video's ensemble performance. All reroutes in Figure 5.10 feature types are economical, using types already evident in one of the four collections. Because of this, the routes still carry an insular relationship among the sync collections and add a sense of unity among the collections.

The colored blocks show that despite the initial fluctuations among the three contextualization schemes, totality takes over from the midsection (Sp.23) until the end, even supporting the metric [M] type sync at the end (indicated straight lines). The outlining of contextualization schemes can also suggest general a general form for the sync trains' emergent rhythmic profile. While a straightforward ABA' form is not evident here, the sync train still follows a general trajectory of increasing action through reroutes towards the midsection and then coming to a figurative slow at the end with the M[P]-[I^2] sync. The sync train suggests a strong connection between *totality* and [I^2] type sync, even to the point that any reroute involving [I^2] has a green color shade marking *totality*.[154] The connection between totality and [I^2] type sync complements the inherent nature of [I^2] sync that highlights equal markedness of musical and visual accents.

So far, all information is inferred from the sync train without need for a score or video. However, it is often the case that sync trains like that of Figure 5.10 must be constructed through analyzing a music scene or video. Therefore, Table 5.1 provides a play-by-play commentary that details the audiovisual events unique to the Nightcord version even though the chart alone does not dynamize the filmic process in the same manner as sync trains.

[154] [I^2] sync are synchronicities that are iconic and isomorphic. According to the Chapter 2 table, the [I^2] sync occur when "Visual accents emphasize a musical feature both rhythmically (isomorphic) and non-rhythmically (iconic).

The fluctuating context types from Sp. 1–22 rely on details to sort out which type occurs. Figure 5.11 excerpts the beginning of the music video to explain why visuality is assigned. The blue shading indicates that aurality is salient in the beginning. Aurality is assigned to this sync collection because the imaginary "camera" zooms out from the fountain (Sp.1–4). The music, transcribed below the sync train, illustrates how the R[D]-[ISO] emphasizes the purely durational aspect of the melodic line as the f#m: i6-VI7-III-VII axis progression. And as mentioned in the prior Figure 5.8, the musical markedness is more salient than the number of marked visual accents, *aurality* is assigned as the main type.

Figure 5.12 excerpts the most problematic area (Sp. 5–16) featuring numerous reroutes and fluctuating markedness types. Paired with wavy lines indicating [R] sync, multiple wedging lines chart out the uncertainty that viewers would experience in real-time and without the foreshadowing provided by the train graph. This sync collection features both prominent onscreen choreography and [VCO] shot changes to emphasize who is singing at a given moment. From the events outlined in the with the play-by-play table, the visuality annotated on the sync train plays into the nature of dance and animated choreography, since visual markedness is more prominent than musical accents. For example, Sp.12–17 features four shot changes focused on the main character Ena, before moving on to another character named Kanade. During this span, the shot changes even cut across the musical phrasing (Sp.13–15) found in the transcription to focus on character expressions. The frequent wedges reroute from visuality to totality, causing two competing levels of contextualization schemes to arise. Nevertheless, visuality remains prevalent because the reroutes do not meet the tripartite requirements to establish or project a new sync collection.

Totality takes over in the mid-section of the sync train (Sp. 23–46), paired with the [I^2] sync type. Increased sync point frequency and reroutes indicated in Figure 5.13 intensify the *audiovisual rhythm*, dynamizing the sequence. Here, the sync train calls attention to a peculiar relationship between increased action and decreased timespans. For example, the number of sync points from SP.36–41 is greater than the one-second timespan that encompasses them. The transcription of this moment highlights how there is a shot change on every note linked to the lyrics singing "bye, bye, bye, bye." Ironically, the changes are not [VCO] because the whole group is singing in that span, the shot changes function more for dramatic emphasis than sonic fidelity. This irony plays into the unpredictable nature of this passage, represented by wavy lines.

Figure 5.14 demonstrates another function of contextualization schemes: mending together different sync collections. The figure excerpts the M[P]-[I^2] sync correlating to the precise moment the black and white animation is colored. Straight lines and regularity (metric periodicity) contrast the prior non-periodic monochromatism with colorful stability. Although the stretch of Sp.47–55 brings in new elements of regularity and metric projection signified by straight lines, this section is still unified by the blanketing totality, shown through green highlights. And because totality is preestablished in the R[S]-[I^2] sync collection, the absences of wedging lines and differences of regularity (metric and straight lines) versus irregularity (rhythmic and wavy lines) is mitigated. Keeping totality as the common denominator among the parts, the contrast of periodicity provides a rhythmic "slow" and steadiness towards the end of the sync train. Therefore, the rhythmic profile that is intensified by prevalent [R] sync from Sp.5–46 is given a deceleration and winding down in the last sync collection (Sp.47–55).

Typical to the *Colorful Stage* 3D music videos, the Nightcord version of "Close to Gray" adapts only a part of the full song available on their album release. Even so, a loose formal

organization is possible, even if the audiovisual form of Nightcord's "Close to Gray" does not tie into an ABA' format like the prior *Nevsky* analysis. According to the layout of context schemes, there is a linear drive towards totality: aurality-visuality-totality-totality. Although ABCD is a straightforward organization of the sync train, I posit a binary AB organization between the *non-totality* (i.e., aurality and visuality) and *totality* sections. Sync points 1–22 precede the chorus of the song that is matched to Sp. 23. Totality contextualizes the chorus and dynamic ending when the animation is colored at Sp. 47. This two-part form reflects the audiovisual interactions, the meta-narrative of Ena's character arc, and the visually contrasted duality between monochromatism and colorfulness.

Contrasting Nightcord's choreographed version, the second adaptation is a Vocaloid version that adapts the entire song and tells an abstract story through animation. Figure 5.15 maps the first half of the Vocaloid version for purposes of a direct comparison with the Nightcord version. Like the Nightcord version, the Vocaloid version's sync train can be constructed purely through watching the music video and reverse-storyboarding the sync points according to timestamps.

The animation for the Vocaloid version, produced by Tsuda (津田), is metaphorical and open for interpretation. One such interpretation is that the main character (hereafter, MC) is struggling with his work as a music producer. The cognitive dissonance formed from his work stress and feelings of being "talentless" is embodied by his doppelganger, or "other self," that he meets in dream sequences. The frequent black dots in Figure 5.15 suggest that the audiovisual narrative relies more on shot changes or between-shot rhythms than in-shot action. There is much less in-shot action and many of the effects, like that of Sp. 46 and 55, are swipe transitions that intensify shot changes. Table 5.2 provides a companion play-by-play to Figure 5.15's sync train.

The commentary shows that most of the audiovisual accents are more action-oriented with few [VCO] relations. This characteristic complements the role of music as a non-diegetic narrative device and contrasts with the diegetic "performance" of the choreography version. Instances of aurality (blue shades) back this claim, as this means that the images of in-between spaces are not as marked as it would be for a choreography video.

Figure 5.15 contains six sync collections, compared to the four collections in the Nightcord version. This suggests that Figure 5.15's rhythmic profile is more dynamic than the choreography version. Freed from focuses on performance and audio fidelity, the Vocaloid version features more durational emphases per the R[D] type sync. As opposed to [Sonic] or [Phrasal] types, the emphasis on duration [D] means that the onscreen action can be limited or abstract if the sync points emphasize the same durational values. This idea is supported by the lack of visuality in this example (yellow shading). The absence of straight lines and prevalence of wavy and wedging lines indicate a highly unstable, intense, and dynamic audiovisual experience. This contrasts with the Nightcord version that moves from irregularity towards regularity at the closing.

The sync train reveals that the Vocaloid adaptation switches between two contextualization schemes: aurality and totality. The abundant shot changes and lack of in-shot motion prioritize *aurality* as the main form of markedness. Interspersed *totality* usually contextualizes emphatic lyrics in areas like Sp. 9–12 or the raven's appearance at Sp. 27–30. Only in the last stretch of Sp. 45–58 does totality help unite audiovisual information and connect sync points that drive the rhythmic profile into intensifying the climactic showdown between the MC and "other self." This intensification is visually represented by irregular wavy (non-periodicity) and wedging (rerouting) lines and the lack of metric periodicity.

For both Nightcord and Vocaloid versions, there is a relationship between totality and [I^2] sync. Outlined in the Vocaloid version's sync train, almost all totality-oriented pockets and stretches coincide with a form of [I^2] visual accent. In contrast, aurality underscores all [ISO] areas, or moments that are not [I^2]. The [I^2] accent is a simultaneity of isographic and iconic relationships; therefore, attributing aurality to a *singular* relation [ISO] clarifies that aurality is a non-simultaneous form of sync markedness. Both maps highlight this consistent relation, with the Nightcord sync train confirming this with visuality connecting singular [ICO] and [VCO].

Rhythmic intensification is also propelled by the increased types of [R] sync collections. The increased types over a short time span create the illusion of "motion" between the frames, thus giving more function to the in-between spaces. In this case, the imagined line *i* is created through stitching shots together in contrast to the Nightcord version's constant onscreen motion. Based on Table 5.2's event timelines, the train features more sync type varieties, shot cuts, and screen effects, despite having less in-shot choreography and action than the Nightcord version. For example, the span of Sp. 27–30 centers on a black raven entering the screen while the main character covers his ears. These four shots are intensified by frequent cutting rhythm that dramatize the entrance of the raven at Sp. 28.

The few [VCO] moments are peculiar because there is only expression of singing, but no sonic fidelity. At Sp. 26, the MC (main character) appears to mouth the lyrics intensifying the lack of fidelity by borrowing the disembodied Vocaloid voicebank of Kagamine Len. Even when the lyrics are synchronized to music and shot changes, there is no emphasis on who is singing. For example, the R[D]-[I^2] that initiates Sp. 9–17 seem to match the *duration* of the music more so than the lyrics "denpa" (digital), "sekai" (world), and "rainy" that happen to coincide at the same time.

Finally, the Figure 5.15 sync train presents a problem not found in the Nightcord version: a *double-reroute* at Sp. 13–15. Here, Sp. 13 and 14 reroute to R[D]-[ISO] for a medium shot of the MC at the computer (Sp. 13) and a mouse cursor moving across the screen (Sp. 14). Instead of returning to the salient R[D]-[I^2] train, the R[D]-[ISO] reroutes again to an R[D]-[ICO], marking three reroute dots on the map. Since the three reroutes are *not all the same type*, the tripartite preference is not met and the R[D]-[I^2] maintains its status as the salient sync type in this span.[155] However, reanalysis is necessary if any double reroute exceeded three total synchronous "beats" (represented by dots on the graph).

Among the many details elucidated by constructing sync trains, this chapter's comparison of "Close to Gray" highlights how the Nightcord version drives action through constant [ICO], [ISO], [VCO], and [I^2] correlations with the music or lyrics. Because the adaptation is oversaturated with character and camera motion, the shot changes and cutting rhythm help viewers identify [VCO] sonic fidelity and important choreography points. The Vocaloid version features no choreography, less in-shot motion, and no vococentricism. Therefore, cutting rhythm is used to intensify the audiovisual narrative and create illusions of motion, such as the entrance of the raven. Through shot changes, the raven appears to "enter" the shot behind the MC character. This chapter's case studies clarify that sync trains are ultimately not meant to reduce or separate a film into component parts. Instead, it is meant to read between the lines and discover the emergent meanings formed by synchronous relationships, which is but one of the many relations formed through a filmic experience.

[155] This preference is defined in Chapter 4 as the accentuation of "at minimum three equidistant occurrences; if not, they maintain status as lower-level, rhythmic sync points."

5.4: Connecting the Dots

Constructing sync trains clarifies the rhythmic profile of animated musical sequences or videos. The train addresses the premise of contextualizing sync points: by adapting Lewin's line *i* and plugging in sync point analysis into the imaginary line *i*. Then the functions of periodicity and presentation (contextualization) define the spaces between each sync point. The map provides data on the frequency of sync point types, prevalent sync collections, regularity (metric periodicity), irregularity (non-metric periodicity), competing sync types, and more. Sync trains' purpose is to provide a generalized map of functions for *both* the individual points and the undefined, contextual spaces that exist between any pair of sync points. The functions are defined by their regular/irregular periodicity and contextualization scheme, or markedness.

By its nature, sync trains are somewhat contradictory relating synchronicity and mickey-mousing to Eisenstein's film editing techniques that predated animated camera and character motion. Yet as exemplified by the very sync trains, Eisenstein's rhythmic editing can serve as basis for identifying sync points *connected* by an imaginary line. That line is the emergent *audiovisual rhythm* that is *propelled* by periodicity and fluctuations among the three contextualization schemes (totality, visuality, and aurality). These interactions mirror the dynamic film process and demonstrate how periodicity and context can connect Eisenstein's "shots A and B" by adapting Lewin's imaginary "line *i*" and defining it with sync train methodology.

Sync trains mark predominant sync techniques, arrange sync collections into a formal hierarchy, and dynamize otherwise static sync points. By connecting the synchronous dots, sync trains illustrate fabricated movement. As Pearlman explains, film rhythmic movement "is the real concern of editing theory and philosophy," and "rhythm in the arts, as in the sciences, is

understood to be patterned movement over time."[156] Sync trains are constellations of patterned movements that are marked by synchronicity. That is why reroutes and wedging lines affect what Pearlman would call "trajectory phrasing." Functionally similar to Hasty's metric projections, trajectory phrasing describes "the manipulation of energy in the creation of rhythm" through cycles of tension and release.[157] Sync trains elucidate buildups of rhythmic tension through tracking rates of reroutes as well as changes in sync types and collections. Sync trains also track moments of "release," or more so stability, in areas of metric sync or where projected sync types are reinforced over a span of different timestamps. Ultimately, the *rates* of the emergent third rhythm (encapsulated by the sync train) dramatize narrative tension and release.

Next chapter, rates of synchronicity, tension, and release are taken to another level. Animated music videos like "Close to Gray" create even more emergent rhythms once added to a rhythm game, where players interact with music, image, and rhythm through mediums such as mobile devices or game controllers. This extra dimension of synchronicity transforms observers into participants, where players negotiate various audio streams according to Stephen McAdams and Albert Bregman.[158] To complete a song, players (sub)conciously evoke Christopher Hasty's metric projection and Justin London's metric entrainment, navigating sync point types in real time. Through touch-based (e.g., haptic) actions, the sync train for "Close to Gray" will intertwine with an extra rhythmic stream that combines animation and real-world sense to realize synchronicities of a higher order.

---

[156] Pearlman, 149.
[157] Ibid., 155.
[158] Stephen McAdams and Albert Bregman, "Hearing Musical Streams," in Computer Music Journal 3, no. 4 (1979): 26–60, http://www.jstor.org/stable/4617866. McAdams and Bregman discuss their concept of an "auditory stream," is a psychological organization of sounds, including music, according to sonic or music contexts in which "sound is embedded." I posit that rhythm games requires active and immediate organization of sounds to contextualize synchronous tap notes to a fixed audio track and complete a song stage.

Chapter 6

Beyond Mickey-Mousing: Rhythm Games and Haptic Synchronicities

Sync point analysis does not need to be limited to dynamic filmic experiences. This chapter pushes sync train analysis beyond film media and applies the framework to mobile rhythm games that use animated objects to design and control audiences' experiences of synchronicity. Contrasting the passive filmic experience, mobile rhythm games offer a haptic, or touch-based, relationship between audience-turned-participants and mobile devices. These active touch-based interactions are guided by fixed rhythmic patterns called "beatmaps." Although beatmaps may closely mirror the music soundtrack, they are their own independent rhythmic streams. Therefore, this chapter's case studies argue that tactile patterns in rhythm games serve as useful analytical hypertext for sync point analysis. Constructing sync trains for analysis additionally demonstrates that Justin London's theory of metric entrainment and Christopher Hasty's theory of metric projection are effective frameworks that can model and mirror players' real-time interaction with beatmaps.

In this chapter, I will show how my sync point and sync train methodologies can extend to the relatively untouched beatmaps of mobile rhythm games. This involves adding another sync train level (or "track") to represent a third rhythmic profile, and then examining *three-way sync* among the music, image, and rhythm game (beatmap) elements. Any three-way sync will represent a "junction" or meeting point that may form a new synchronous relationship.

Using sync trains as an analytical framework to track three-way sync applies music theories of rhythm and meter to understand gameplay, even if gamers are unaware of said theories. After producing data through analysis, a focal case study of "Close to Gray" shows how sync trains elucidate tripartite synchronicity across three independent and interacting rhythmic

streams: music, image, and gameplay (beatmaps). To define said tripartite synchronicity, this chapter is divided into three sections that explore (1) the "player-driven" sync between music and game, (2) how music bridges synchronicities between gameplay and background animated videos, and (3) how trilateral sync redefines functions and definitions of sync labels n[x]-[y].[159] All related case studies aim to push the boundaries of the sync point analytical framework to deal with both the rise of mixed-media animations and continual technological advancements beyond traditional animated films.

6.1: Background: Player-driven synchronicity

Separate from audiovisual rhythm, the beatmap is an aggregate term for multiple elements involved in providing touch-based (haptic) sync points. Julianne Grasso defines rhythm games as "fundamentally linked with gameplay," being both "task-based" and "task-triggered."[160] Karen Collins attributes rhythm-action game's task-oriented goals to a sense of musical embodiment and virtual performance, even if the sound itself is separated from its original source and applied to virtual representations of musical instruments.[161] Therefore, for the purposes of this chapter, rhythm games are treated as a subset of video games centered on players tapping virtual objects in time with music to fulfill a rhythmic sequence.

Rhythm gameplay requires a controller or media device for players to input rhythmic combinations, discussed later in this chapter. This real-time action makes games more immediate

---

[159] Detailed in Chapters 2–4, the n[x]-[y] labelling system assigns function to sync points. The n[x]-[y] formula is a placeholder for the following: *n* = rate of occurrence; [x] = music accent type; [y] = visual accent type.
[160] Julianne Grasso, "Video Game Music, Meaning, and the Possibilities of Play," PhD. dissertation, (University of Chicago, 2020), 29. Grasso mentions *Rhythm Heaven*, *Guitar Hero*, and *Dance Dance Revolution* as examples of games where players must interact with game elements that have direct effect on their musical experiences.
[161] Karen Collins, "Embodying Game Sound in Performance: The Real and the Virtual," in *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games* (Cambridge, Massachusetts: MIT Press, 2013), 19. She writes that "Game sound, in this sense, becomes both a tool for and a site of performance and spectatorship." (p. 91).

and interactive than passive film watching. In his pedagogy-oriented article featuring *Dance Dance Revolution* (*DDR*), Brent Auerbach describes *DDR* and related *Bemani* games as focused "on the execution of rhythmic, as opposed to pitch-based, skills.[162] Auerbach explains the term *Bemani* as an abbreviation of "Beat Mania," and an umbrella term for a family of games that use human body-related actions to interact with music.[163] Auerbach not only makes the case for *DDR* as a useful computer assisted instruction (CAI) device, but also a gateway into eurhythmics due to *DDR*'s (and by extension rhythm games') high levels of interactivity and ability to engage an active brain.[164] My chapter expands on Auerbach's ideas with the claim that mobile rhythm games are as similarly highly interactive and engaging as DDR, and by virtue of being—as Julianne Grasso would describe—a task-oriented game genre. Yet, according to Peter Shultz, sometimes rhythm games such as his case study on *Rhythm Heaven* "requires an environment in which actions happen somewhat predictably, at regular intervals."[165]

Once a player engages with the game, they (un)consciously negotiate various audio streams. The most immediate are the generalized streams of the music soundtrack and an independent rhythmic stream called a beatmap. A game "stage" comprises an audio track and the rhythmic beatmap mapped over the song track. This chapter's definition of *beatmap* refers to a fixed rhythmic stream that mirrors Stephen McAdams and Albert Bregman's definition of "auditory stream," or a listener's psychological organization of music, sound, and sonic

---

[162] Brent Auerbach, "Pedagogical Applications of the Video Game *Dance Dance Revolution* to Aural Skills Instruction," in *Music Theory Online* 16, 1 (2010), accessed September 29, 2023, https://mtosmt.org/issues/mto.10.16.1/mto.10.16.1.auerbach.html, para. 5.
[163] Ibid., para. 4. Auerbach mentions games like *Dance Dance Revolution*, *Guitar Hero*, and *Rock Band* as effectively engaging players with rhythm through use of physical (haptic) input of some kind in accordance with non-verbal, protonotation-like animations of rhythmic combinations.
[164] Ibid., para. 13 and 14.
[165] Peter Shultz, "Rhythm Sense: Modality and Enactive Perception in *Rhythm Heaven*," in *Music Video Games: Performance, Politics, and Play* (New York: Bloomsbury Academic, 2016): 269.

contexts.[166] I argue that while auditory streams are complex phenomena, sync points help orient listeners to specific musical accents. I apply the same idea to rhythm games, where touch-based, onscreen objects stitch together separate elements of music, gameplay, and accompanying animated films. Daniel O'Meara's chapter on the rhythm game *Rocksmith* also points out that the design of like rhythm games often "prioritize the beginning of phrases," yet can also present irregular phrase groupings that might be "odd from the standpoint of other musical parameters…"[167] This holds true for this chapter's case study, although I seek to demonstrate how sync trains can organize and resolve any such potential conflicts.

While different rhythm games have their own mechanics for rhythmic input, all mobile rhythm games rely on some form of touch-based sync with an interface that hosts the beatmap. This chapter's case studies revolve around a gameplay interface that features three general types of onscreen action typical to mobile rhythm games:

- The input zone (a touched-based/tactile area)
- Beat tiles (animated rhythmic objects)
- Background image/video

Figure 6.1a snapshots the interface of the mobile rhythm game, *Hatsune Miku: Colorful Stage!*. The semi-transparent bar at the bottom of the screen serves as an input zone where players tap *beat tiles*, or animated objects appearing in real-time.[168] The beat tiles (like those circled in Fig.

[166] Stephen McAdams and Albert Bregman, "Hearing Musical Streams," in Computer Music Journal 3, no. 4 (1979): 26–60, http://www.jstor.org/stable/4617866.
[167] Daniel O'Meara, "Rocksmith and the Shaping of Player Experience," in *Music Video Games: Performance, Politics, and Play* (New York: Bloomsbury Academic, 2016): 239–40.
[168] The beat tiles can be understood as a form of *protonotation*, like that found in Karpinski's. This is because the tiles are generalized representation of rhythm and pitch, with no immediate necessity for music literacy. For more, see Gary Karpinski, *Aural Skills Acquisition: The development of listening, reading, and performing skills in college-level musicians*, (New York: Oxford University Press, 2000).

6.1a) move *towards* the input zone and must be tapped the exact moment they reach the input zone. The interface of the input zone and beat tiles are superimposed over a background. In *Colorful Stage!*, the background is either a static picture of an album cover or related animated video. The tiles form the beatmap's independent rhythmic stream but may synchronize with any of the other music or image rhythmic streams.

Beatmap analysis is complex because synchronicities are achieved through an aggregate of touch-based taps (beatmap) that may involve both hands, onscreen markers (beat tiles), and a correlating song (audio). Although the beatmaps are fixed, players' accuracy and tapping the onscreen animations determine their individualized experiences of rhythm, mistakes included. Players use correct hand positions to tap synchronously and accent parts of the music soundtrack or background video. The game's grading system evaluates players' accuracy in synchronizing beat tiles to the music. From worst to best, players can earn a "miss," "bad," "good," "great," or "perfect" for every beat tile they tap according to the fixed maps. The total score earned from correct rhythmic input can earn score ranks ascending from C, B, A, and S score evaluations, with the most optimal grade earning a "full combo" achievement.

While the tiles often have their own sound effects to gauge player accuracy, these sounds are omitted from analysis. Instead, the analytical focus is on the fact that sound effects exist to mark an intentional synchronous relation between music and beatmap. This idea is represented in Figure 6.1b, where tap notes are transcribed onto sync rhythm line. Illustrated in Figure 6.1b, beat tiles in rhythm games like *Colorful Stage!* comes in three basic varieties:

- Tap notes
- Flick or swipe notes
- Hold and/or slide notes

The annotated "tap notes" in Figure 6.1b are an instant rhythmic action where players input tap instantaneous and short durations through a tapping motion. These notes have a 1:1 ratio or note-to-note sync relation between tapped beat tiles and a musical accent. This relationship forms a rhythmic profile between music and gameplay that is *separate* and different from the filmic synchronicities covered in Chapters 2–5.

The flick or swipe note is a tap note with additional action such as flicking or swiping (forward or sideways) on the mobile device screen. The red tile with a triangle in Figure 6.1b (numbered 3) indicates flick/swipe notes. The action is instantaneous, much like the tap notes. As seen in the transcription (numbers 1 and 4), tap notes can also occur simultaneously, with hands or fingers tapping multiple tiles at the same time. The varieties of tap, flick, and swipe notes are all inherently *isomorphic* or [ISO], especially since they mirror the durational structure of the notes they sync to.[169] Through gameplay, this [ISO] relation is dynamized through synchronized, haptic human touch with animated onscreen objects. The right-side transcription in Figure 6.1b matches the order number (1, 2, 3, 4) to the transcribed rhythm. The right-hand order sequence would be played as tap/ tap/ flick/ tap. Simultaneously, the left-hand order would be played as hold/-/ release/ tap.

The last beat tile type, the hold note, has the longest durational value. It is held after an initial tap and involves three physical motions: tap, hold, and release. Hold notes broadly encompass non-rhythmic contours and musical pitch strains or phrases. In relation to Chapter 2's visual accent types, hold notes that exceed the 1:1 sync ratio creates *iconic* [ICO] relations with sounds of group cheers, non-musical sound effects, cymbal crashes, electronic effects, and so

[169] Covered in Chapter 2, [ISO] indicates a visual accent that focuses on temporal structure and mirrors the durational value of the music.

forth.[170] Hold notes that exceed the 1:1 sync ratio and involves pitch strains are [I^2] and emphasize a musical feature rhythmically and non-rhythmically.[171]

Hold notes problematize analysis because they include all musical materials that coincide with the physical action tap, hold, and release. These actions are physical motions and do not need to mirror phrasal motions found in the music. Figure 6.2 features transcribed tap rhythms synchronized to a musical excerpt. The rhythm of "Close to Gray" is swung, as indicated in Figure 6.2. This instance demonstrates how beatmap tiles can mirror swing rhythms without visually indicating it. Both right- and left-hand parts are condensed into one rhythmic line with note stems up representing the right hand and stems down representing the left. The annotated [I^2] refers to the initial tap and holding action during gameplay. In rhythm games, the "release" action may come with an additional sound effect to aurally mark the end of a hold note. For the sake of analysis, the hold note is analyzed for its initial tap and the release marker is counted as part of the [ICO] relation instead of an additional rhythmic [ISO] "tap" or sonic mark. Long note durations, ties, and a circle-cross notehead in Figure 6.2 mark the "release" action of the hold note. The notational practice of marking the "release" of a hold note adapts Cheuk Ling Yu's use of x-shaped notation heads for a beatmap transcription from the anime idol rhythm game *Love Live! School Idol Festival*.[172] However, notating the release action as well as using ties for hold notes differ from Yu's original system.

---

[170] Also covered in Chapter 2, [ICO] describes a sync point function whereby the visual accent [ICO] reflects the non-rhythmic contours of music. This may include timbre, pitch, melodies, fragments, motives, etc.

[171] The [I^2] function refers to a visual accent that simultaneously emphasizes both the rhythmic [ISO] and non-rhythmic [ICO] aspects of the music.

[172] Cheuk Ling Yu, "Japanese Anime Idol Music in the Media Mix: A Case Study of the Love Live Idols," Master's thesis, (University of San Diego, 2021), Retrieved from ProQuest, https://uc.idm.oclc.org/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fdissertations-theses%2Fjapanese-anime-idol-music-media-mix-case-study-em%2Fdocview%2F2625044103%2Fse-2%3Faccountid%3D2909.

Yu's system is more sophisticated, numbering protonotation tiles for each of the nine haptic inputs for the game's specific interface. However, Yu does not indicate any sort of flick or swipe notes in the transcription. Yu also notates two voices for simultaneous taps, later referenced in this chapter.

Rhythm games' interface and control help design engagement and immersion for gameplayers. Sync points are especially critical to rhythm games because the genre evokes what Kiri Miller identifies as "schizophonia," or formed cognitive dissonance between a recorded song featured in gameplay and its source emanating from a mobile controller, as opposed to maintaining fidelity to the source instrument or voice.[173] The disembodied sounds, compounded with beatmap tiles, add to what Miller suggests as rhythm games' ability to "dramatize schizophonia, endorsing the idea of a split between the live and the recorded and inviting people to play at mending that split."[174] I argue that sync points in rhythm games function as player-driven seams that mend the "schizophonic" split, creating marked and objective points for gameplay and analysis. The mending of splits places synchronicity in the hands of players, creating interactivity that moves beyond passively experiencing filmic synchronicities.

Rhythm games heavily focus on player-oriented action, or "player-driven" experiences. Correct rhythmic inputs allow players to experience the complete song from beginning to end, while failure ends the song prematurely. Because of the inherent need for correct input, I argue that the "player-driven" aspect of rhythm games according to Noah Kellman's definitions, underscores touch-based synchronicity.

Noah Kellman's concept of "player-driven" gameplay is where "Player choice, game state, and music all have a direct, reciprocal effect on each other."[175] I adapt Kellman's original

---

[173] Kiri Miller, *Playing Along: Digital Games YouTube and Virtual Performance* (Oxford: Oxford University Press, 2012). Additionally, in her chapter "Interacting with Sound: A Theory of Action, Image, and Sound," Karen Collins credits R. Murray Schafer for the term "schizophonic," as meaning a split of sound from its source. The sounds become disembodied but then are embodied by players through gameplay. For more, see Karen Collins, "Interacting with Sound: A Theory of Action, Image, and Sound," in *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games* (Cambridge, Massachusetts: MIT Press, 2013), 19.
[174] Miller, 86.
[175] Noah Kellman, "Designing Interaction between the Player and the Music," in *The Game Music Handbook: A Practical Guide to Crafting an Unforgettable Musical Soundscape*, (New York, 2020; online ed., Oxford Academic, Sep. 23, 2021), 105, https://doi.org/10.1093/oso/9780190938680.001.0001, accessed 21 Aug. 2023.

ideas to the priorities of sync point analysis: *haptic sync* is a method of "player choice," *synchronous state* is the optimal rhythm "game state," and the *generalized audio stream* serves as "music."[176] These relations form through player engagement and intensify when players enter a synchronous state entrained by touch-based actions.

Player interactivity and the sync points they encounter are based on the difficulty of a chosen game stage. Difficulty levels play a large role in determining the frequency and types of [ICO] or [ISO] sync emergent from the beatmap. *Colorful Stage!* offers five difficulty levels in ascending order: Easy, Normal, Hard, Expert, and Master. Figure 6.3 transcribes Normal, Hard, and Expert levels to detail the differences. At the Normal difficulty level, the tap notes are sparse to accommodate beginner players and are mostly [ISO] and [ICO]. The sparse protonotation tiles align with the harmonic changes in the musical backtrack, with little to no sync to the vocal part. The [ICO] type holds notes mirror the synth sounds within the backtrack. At Hard difficulty, what were [ICO] are now [ISO] and [I^2] taps that mirror the instrumental melodies and vocals instead of the slower synth sounds. The result is a more immediate synchronous relationship between tap tiles and the constantly moving musical accents.

Expert difficulty is more sophisticated, with added difficulty such as using two hands and simultaneous combinations of tap, flick/swipe, and hold notes. For example, Figure 6.3's expert-level sync line features simultaneous tap and hold-notes at the highlighted annotation. Here, the tap notes reflect the [ISO] nature of tap tile while hold notes feature [I^2] sync to emphasize note duration *and* the ensemble's lilting singing and phrasing. Again, simultaneous notes are represented by two voices on a singular sync line, distinguished by upwards and downwards note stems. This method, featured in Cheuk Ling Yu's writings, defines any moment where there are

[176] Kellman, ibid.

two or more protonotation tiles in various styles.[177] This example adapts a similar idea, with additional note heads for different protonotation types.

Rhythm games create synchronous experiences through gameplay design (interface and beatmap), touch-based interactivity, and player-drivenness. Constant onscreen and audio stimuli at the Expert level force players to be responsive and reactive to incoming tiles. Therefore, players who are attuned to the music and gameplay are pushed into a *synchronous state*. The more synchronous they are to the music and gameplay, the more "correct" players' input and high scores are. It is this synchronous state that incorporates the eclectic concepts of player-drivenness, mends Miller's schizophonic lines, and creates a sense of organic, haptic synchronicity. The next section seeks to describe the synchronous state by mirroring and reverse-engineering through constructing sync trains that touch upon the music theories of metric entrainment and projection.

6.2: Methodology: Beatmap Analysis

By nature, rhythm games rely most on correlations with the audio stream to ensure that players' experiences reflect the musical features to a great capacity, as opposed to creating abstracted patterns that evoke metric and rhythmic dissonance. The following section formalizes a methodology for analyzing beatmaps and incorporating its analytical data into constructing sync trains. The application of sync train analysis *generalizes* the beatmap just as the methodology of Chapters 1–4 generalized audiovisual accents to focus on detailing rhythmic profiles. Sync point analysis of beatmaps is functionally similar to constructing sync trains: once a pattern is derived from emergent rhythms, models of London's metric entrainment and Hasty's

[177] Ibid.

metric projection are mappable onto beatmap transcriptions. Incorporating beatmaps into analysis elevates the status of mobile rhythm games and treats beatmaps' rhythmic as analytical "hypertext," as respectively suggested by Jesper Kaae and K.J. Donnelly.[178] In his chapter on the "the triple lock of sound and image to player input," Donnelly specifically relates synchronization to gameplay interactivity, and underlines the importance of player-drivenness in gameplay:

> Synchronization holds together a unity of audio and visual, and their combination is added to player input. This is absolutely crucial to the process of immersion through holding together the illusion of sound and vision unity, as well as the player's connection with that amalgamation.[179]

Rhythm games are straightforward in that beatmaps are immediate and non-narrative. That is, the beatmaps do not have story quests or a narrative built into their rhythmic design, per se. Aside from the possibility of navigating story requirements to unlock songs, the gameplay and immersion is self-contained within standalone songs and associated rhythmic patterning.[180]

To limit the scope of this chapter, the sync point framework is applied only to haptic rhythm games and not to other gameplay designs such as open-world exploration or fighting games that bear their own sonic and rhythmic profiles.

---

[178] Jesper Kaae, "Theoretical Approaches to Composing Dynamic Music for Video Games," in *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*, ed. Karen Collins, (Aldershot: Ashgate, 2008), 77. See also, K.J. Donnelly, "The Triple Lock of Synchronization," in *The Cambridge Companion to Video Game Music*, eds. Melanie Fritsch and Tim Summers (Cambridge: Cambridge University Press, 2021), 94–109, doi:10.1017/9781108670289.008.

[179] Donnelly, 95.

[180] Some rhythm games, including *Colorful Stage!*, may require reading through character stories or other game mechanics to unlock certain songs. However, this fact does not affect the fact that there are beatmaps available for analysis, even with the song list provided at the start of the game.

Analyzing beatmaps is separate from player skill. The maps are analyzed at their full potential, meaning that all protonotation tiles are present for a given stage as if the player achieved all "perfects" and the "full combo" evaluation. Sync points allow for analyzing the complete beatmap as an objective hypertext that rhythm game players individually "perform" and experience to create emergent meanings. The auto-play mechanic in *Colorful Stage!* allows players to observe gameplay and study the beatmap as the program plays the proper and accurate rhythms and protonotation tile combinations, hence the analytical utility of the game.

The Figure 6.4 analysis revisits the opening passage of "Close to Gray" for three levels: normal, hard, and expert.[181] The audio stream represents the generalized music line, condensed into one clef and staff. The levels normal, hard, and expert charts the rhythms for each respective level. Figure 4's annotations bracket recurring moments where three levels consistently sync with the downbeat of each measure. These groupings form hierarchical metric sync between beatmap and music.

Figure 6.4 exemplifies how rhythm games are idiosyncratic, with the easier levels being *less synchronous* than the harder levels that exhibit extensive mickey-mousing. Of the three levels in the example, the Normal difficulty features minimal synchronicity with the audio stream. Aside from syncing the downbeat, the second release note (marked by an x-circle notehead) aligns with the rest on the musical beat 3 instead of the instrumental's sounding pitches. But because the hold note is made of three motions (tap, hold, release), the initial tap is considered the marked sync point. Therefore, the release motion over the rest that is *connected* to the initial tap is considered part of the downbeat, highlighted in Figure 6.4.

[181] Example of Expert-level gameplay and media reference: かんぱり,【プロジェクトセカイ】 限りなく灰色へ EXPERT All Perfect, YouTube Video, 2:20. https://www.youtube.com/watch?v=Px-LtgKICnY.

Although the release note is considered part of the hold note, analysis is still complicated by the fact that the release note is marked by its own sound effect, beat tile, and physical motion. All these forms of markedness set up a real-time expectation that the upcoming beat tile would still align with a note from the salient melody transcribed in Figure 6.5. However, as illustrated in Figure 6.5, the red arrow marks a thwarted expectation if a player immediately expects the second note to match any sounding pitch. It may also thwart a theoretical expectation for there to be a sound on the third beat of Figure 6.5, m. 2. Because of these inconsistencies, projecting synchronicity at the two-bar, hypermetric level can override any micro-level inconsistencies. The topmost blue arrows in Figure 6.5 calls attention to emergent recurring patterns for mm. 2–3, mm. 4–5, mm. 6–7, and mm. 8–9. These units form a sync train unit of HM[P]-[I^2], whereby the downbeat protonotation tiles indicate hypermetric sync HM and the [I^2] indicate how the hold/slide and release notes stand in for a collection of musical pitches from the audio stream. The HM[P]-[I^2] also demonstrates how the beatmap reorients and reorganizes the projection and rhythmic understanding of the song. At the hypermetric level, the downbeat (and not the pickup) is synchronized to facilitate groupings of beatmap patterns in relation to the music.

The Expert level adds another rhythmic layer, signified by notes with downwards stems. To facilitate reading, the upwards and downwards stems indicate which taps are optimally tapped with right and left hands, even though players may compensate for the physical motions with different hand choreographies. Higher difficulties feature more notes and higher probabilities of synchronicity, and the Hard and Expert levels complicate the straightforward metric downbeat relationship found at Normal difficulty. In addition to outlining metric sync, the beatmap closely "mickey-mouses" the audio stream. Shown in Figure 6.6, the Hard and Expert beatmaps include the pickup notes and create groupings shifted from the barline that cut across metric boundaries

provided by the barlines. This shows how beatmaps differ from prior chapters' audiovisual rhythms: that the haptic elements do not differentiate anacrusis and pickups, nor are there visual accents (i.e. shot cuts, camera motion, character motion) that help align or realign audiovisual rhythm at the downbeats. Therefore, the units annotated in Figure 6 are more akin to Lerdahl and Jackendoff's definition of grouping structure, or the hierarchical organization of nonoverlapping elements of non-metrical accent.[182]

When shifted groupings of the beatmap conflict with musical meter, the haptic rhythm takes hierarchical precedence. In Figure 6.6, the beatmap tunes players into interactive gameplay at the anacrusis through haptic touch, displacing the downbeat feel on notes occurring a beat before the actual downbeat of the music. This creates an "out of phase" situation, described by Lerdahl and Jackendoff as "grouping boundaries [that] cut across the periodicity of the metrical grid."[183] The arrow in Figure 6.6 shows that despite being out of phase, the repeated patterns (or recursive groups) still form and facilitate projection and confirmation of rhythmic (beatmap) and musical (the transcribed music) phrases. The M[P]-[I^2] function that is assigned to these groupings *does not* reflect the music's metric grid but rather the *metric nature* of the projection patterns. This is much like how the functional labels in the previous chapters mirror the emergent rhythm formed between audio and image sync points. Therefore, the M[P]-[I^2] stands for recurring groupings of predictable and metric (equidistant and durational) nature through the definition of audiovisual synchronicity.

---

[182] Fred Lerdahl and Ray Jackendoff, "On the Theory of Grouping and Meter," *The Musical Quarterly* 67, no. 4 (1981): 479–506, http://www.jstor.org/stable/742075.

Lerdahl and Jackendoff differentiate grouping from metric structure, stating that grouping is "hierarchical in a nonoverlapping fashion," and is a "recursive" group of "contiguous elements" (p. 484). Additionally, they state that "*groups do not receive metrical accent*" (p. 494).

[183] Ibid., 499. Brackets, mine.

Even with the out of phase groupings, the rhythmic profile of Figure 6.6 does not change its M[P]-[I^2] reading regardless of the physical action required to interact with the beatmap. The only sense of change occurs when sync collections change and the projection of continued M[P]-[I^2] is denied. This is what happens in Figure 6.7's excerpt, where beatmap switches from synchronizing the instrumentals to Ena's vocals. The red arrows and annotation indicate the point where M[P]-[I^2] changes to a different sync type. Figure 6.7 uses a sync train collection to illustrate how the beatmap's M[P]-[I^2] changes to an R[S]-[VCO] in m. 10. This switch contrasts the M[P]-[I^2] with an R[S] sync collection that indicates non-periodicity and unpredictability. The [VCO] supports the sense of change by emphasizing the vocals instead of the previously synced instrumental musical accents.

The analysis of beatmaps as a hypertext takes for granted the fact that beatmaps are already *generalizations* of the music they correlate to. Figure 6.8a exemplifies this in m. 18. While the Expert-level taps predominantly sync to the vocal part, there are beats that coincide with a rest (similarly to Figure 6.5). These taps further emphasize the syncopated nature of the vocal part, and loosely synchronize with underlying harmonic changes in the synth and drum parts. However, the predominance of vocal sync relations maintains the R[S]-[VCO] sync collection.

The generalizations carry over to the last M[P]-[I^2] section of the song, excerpted in Figure 6.8b. Here, the grouping relations return as in the introduction, subsuming any [R]-type asynchronous parts under the out of phase units. Just like in Figure 6.7, hand-swapping patterns, marked by slurs in upwards/downwards stems, generalizes haptic choreography: that despite changes in hands, the rhythmic value of notes for mm. 46–52 remain *the same*. The annotation in Figure 6.8b features upwards and downwards slur lines to indicate the upper or lower voice

being held during gameplay. The line connecting two highlighted moments in Fig. 6.8b indicates the same pattern, transferred to different hands. Figure 6.8c snapshots the patterns through game screenshots. The taps will therefore sound *the same* regardless of which hand is playing the taps. This understanding reinforces the presence of M[P]-[I^2] in the beatmap's rhythmic profile.

The generalizations inherent to beatmaps and various difficulty levels allow them to easily convert into sync trains. The goal of beatmap-to-music synchronicities is immersing players into the gameplay of *Colorful Stage!* and similar rhythm games; the higher degree of sync aids in both stitching together disembodied sounds and touch-based, haptic choreographies. The following focal case study will connect the already visual- and player-driven beatmaps to the visualities of choreographed music videos that also run in the background of a gameplay session.

6.3: Case Study: Trilateral sync in "Close to Gray"

The sheer amount of audiovisual information in a rhythm game can be overstimulating. Players' heightened interactions with gameplay (as opposed to more inactive film-viewing) are individualized. It is possible that one player prioritizes watching the background music video while another focuses on beatmap tiles. Therefore, this chapter argues that music is the determinate component for both beatmap analysis and a three-way sync analysis of a hypertext involving rhythmic beatmaps, an audio track, and background video image (called 3DMV, the shortform of "3D Music Video"). In doing so, the case studies findings ultimately claim that music and its metric elements can help organize the overstimulating elements into digestible units that reflect how rhythm gamers chunk their information.

Going forward, the three-way synchronicity is called *trilateral sync* (hereafter, tri-sync). It is tri-sync that helps organize and call attention to specific moments in gameplay with

overwhelming stimuli. This idea applies sync point analysis on a broader scale to demonstrate that sync points and rhythmic projection still functionally organize media experiences. Figure 6.9 demonstrates this across four notated staff lines representing the audio, beatmap, audiovisual, and tri-sync streams. The focal case study will demonstrate that tri-sync relations build towards trends of metric organization that facilitate both gameplay and players' anticipation of patterns to occur in metric fashion.

The Expert difficulty level is chosen for the focal case study because of its higher degree of mickey-mousing and synchronicities. The Master level is not chosen because the highest difficulty may detract from both listening to music (an over-reliance of eye-hand coordination) or involve a higher frequency of asynchronous notes outside the scope of this dissertation.

The focal case study also differentiates two sync relations: the *audiovisual sync* highlighted in Chapters 1–4 and the *audio-haptic sync* formed between music (audio stream) and haptic beatmaps. Trilateral sync uses music as the common denominator to create an emergent, dialectic, and new rhythmic pattern. The emergent pattern is a profile that is separate from anything drawn solely from the beatmap, music, or video image. This new and tripartite rhythmic profile, expressed through a sync train, may even alter sync type labels and functions. Simultaneously, trilateral sync defines how otherwise film *observers* (i.e., audiences) transform into media interactors and enactors (i.e., players or gamers). Tri-sync analysis is methodical and consists of three steps that revolve around sync train construction and comparison:

1. Construct a sync train to create rhythmic profile for audiovisual relations (video and music sync).
2. Construct a rhythmic profile for audio-haptic sync (beatmap and audio sync).

3. Compare both streams. Any three-way alignment creates tri-sync. Any contradictions are resolved by (hyper)metric hierarchy, including the principles laid out in Chapter 4.

The following case study will demonstrate three general conclusions regarding trilateral sync: that (1) there are potential contradictions in determining sync types, (2) trilateral sync fluctuates between contradictory and non-contradictory moments, and (3) tri-sync is often metric- or phrase-oriented and can occur when the *rhythmic profiles* are the same—even when rhythmic tile patterns switch between hands during gameplay.

Figure 6.10 highlights the first and immediate challenge of adding a third rhythmic stream to audiovisual analysis: potentially conflicting sync labels. The introductory passage (mm. 1–9) indicates R[D]-[ISO] audiovisual sync between music and image, while an M[P]-[I^2] relation forms haptic sync between music and beatmap. The former label matched sync points to the durational values of a camera zoom and the initial character motions of the animated dance choreography. The latter label emphasizes the metric quality mirrored in both music and beatmap patterning, as well as the metric recurring projections shown in the prior figures. Hierarchical relations help resolve such conflictual cases. As discussed in Chapter 4, metric sync types [M] override all rhythmic [R] types while hierarchically subsuming other sync types under the "hypermetric" label. However, as the "AH (audio-haptic) Groups" line shows in Figure 6.10a, the grouping of consistent and recurring elements are out of phase with the "AV (audiovisual) rhythm" line. This is where tri-sync reorganizes lower-level groupings into a more metric logic.

The "tri-sync" line at the top of Figure 6.10a's transcription handles conflicts between the Audiovisual rhythm and audio-haptic rhythm and grouping structures. Tri-sync's overall rhythm is slower than the immediate ones of the haptic "Expert" and musical "Audio Stream" lines in

the transcription. As a matter of fact, the tri-sync line has one less sync point than the audiovisual rhythm and audio-haptic groupings.[184] This is a result of tri-sync's generalization and omission of any points that do not match across three lines. As a result, Figure 6.10a's tri-sync rhythmic profile features rhythms slower than that of the beatmap, audio, or visual profiles. Figure 6.10b visually summarizes this factor.

Figure 6.10a's tri-sync aligns with the audiovisual rhythm in a metric way that reflects Sergei Eisenstein's definition of metric being mathematically determinable lengths.[185] The mathematically determinable lengths cut through the audio-haptic groupings, creating an M[D]-[I^2] relationship for [Duration] instead of the audio-haptic group's M[P], or [Phrasal], accentuation. This is how the tri-sync line merges the audiovisual rhythm's [Durational] with the audio-haptic group's [Metric] and [I^2] qualities, resulting in the emergent M[D]-[I^2] relationship annotated in Figure 6.10a. Naturally, the tri-sync is not what the player actually experiences in real-time, as that would be the AH Groups rhythm unfolded with tactile interactions of the "Expert" surface level. However, tri-sync is the emergent rhythm where *all* components are conveyed to the player at that exact, synchronous moment. This gives insight into an emergent form created by different streams of synchronicities.

The haptic and immediate nature of rhythm games facilitate reorientation from the audio-haptic groupings to the tri-sync because all elements of music, image, sound, and gameplay funnel attention to the audiovisual rhythm. However, a viable concern arises when tri-sync detracts from forming a narrative, especially with the immediacy of haptic patterns and the fact that gameplay interfaces are "animations" in themselves. Though immediate, haptic features are

[184] In these cases, the sync points numbers still refer to the original AV rhythm line to demonstrate the differences created by tri-sync.
[185] Please refer to Chapter 2 for in-depth definitions.

always organized through metric projection or retrospective realizations of sync trains and patterning. Figure 6.10b's sync train demonstrates how dialectic conflicts between two sync types create new meaning.

Figure 6.10b is the first instance of a sync train that has *multiple* parallel lines, or "tracks." The lower line represents the referential *audiovisual sync* detailed in Chapter 5, while the top line represents the tri-sync train formed in Figure 6.10a.[186] The vertical box enclosing the R[S]-[VCO] node indicates *pivot points* where the tracks *switch* between emphasizing the tri-sync's rhythmic profile and defaulting to the referential audiovisual sync train. Using pivots to essentially "switch" between the sync train tracks illustrates both fluid liveness of gameplay and the salient features that aid in grouping beatmap patterns. For example, the M[P]-[I^2] is a technical way to indicate that the beatmap patterns are grouped according to musical phrases. This reading conflicts with the original sync train's reading of R[D]-[ISO] (bottommost line) and *relocates* the position of the sync points, as indicated by the annotated measure numbers. Likewise, the sync points are again relocated when tri-sync reorganizes the conflicting labels.

The M[P]-[I^2] of the AH Groups emphasizes the beatmap's mirroring of the music as well as the hierarchical and organizational nature of metric sync. What was purely transitional camera motion zooming out from a close-up of the water fountain becomes a *framing visual accent* for both the musical [P], or [Phrasal], accentuation and the recurring beatmap patterning in mm. 1–9. As discussed in prior chapters, the straight and wavy lines in Figure 6.10b illustrate a theoretical projection trajectory. Upon the second iteration of the beatmap patterns, players associate the pattern with the transcribed musical notes and project repetition of said patterns, which recur

[186] The audiovisual reading excerpts the sync train from Figure 10a of Chapter 5.

twice afterwards. While the M[P]-[I^2] is transformed at the tri-sync level, the reroute in Sp. 11–13 still maintains the old label.

The presence of multiple tracks indicates conflict, while switching back to a single line represents conformity. The tripartite convergence at R[S]-[VCO] indicates that all parts work together to emphasize the vococentric accents in the next section. Figure 6.11 demonstrates this common behavior for tri-sync to fluctuate between contradictory and non-contradictory function labels within a sync train. Figure 6.11a excerpts a passage where the Expert-level beatmap reinforces the [VCO] visual accent type. Here, the reroutes at Sp. 11–13 pose an analytical problem: the tri-sync (top) level produces conflicting readings from the other rhythmic levels and includes additional reroutes. One of the rhythmic profiles is removed here because there is no additional conflict in terms of audio-haptic phrase groupings. Therefore, the sync train is simplified to two tracks.

Detailed in the Figure 6.11b sync train, there is a reroute (Sp. 11 and 12) originally labelled R[S]-[I^2] that is reinterpreted to an M[P]-[I^2], even though the original R[S]-[I^2] highlights the immediate synchronization of the "Rainy, rainy," phrase at the *audiovisual* level. The reroute emphasizes the recurring behavior that beatmaps, and tri-sync by extension, tend to reflect the musical phrasing and create (hyper)metric relations. This mirroring between beatmap and music highlights music's role as the "seam" of tri-sync. The addition of a third reroute fundamentally changes Sp.13 but does not feature three R[S]-[I^2] or M[P]-[I^2] sync types in a row that would change the entire R[S]-[VCO] collection to a different one. At the same time, the extended reroute also alters the contextualization type. Although this changes the overall train ever so slightly, the addition of another instance of totality emphasizes how beatmaps tend to rely on all elements (hence [I^2]) to present tap tiles. This characteristic is elevated to the tri-sync

level, where beatmaps' phrasal and metric tendencies *transform* and reinterpret the R-type syncs of the audiovisual track.

The previous figures exemplify how music balances any arising conflicts of function between the visual and haptic streams. Sometimes, there is no conflict at all. This is exemplified in Figure 6.12 that focuses on the closing section of "Close to Gray." Here, the colored beams demonstrate how sync collection generalizes sections that maintain the same rhythmic profile, despite hand-switching. At the end of the song, the beatmap conforms with the M[P]-[I^2] label from the sync train. However, the up- and down-stems in Figure 6.12 indicate that the rhythms in mm. 50–52 are the same, but switches between right and left hands in terms of haptic choreographies. Disregarding the hand switch, the release and flick notes are functionally the same, occurring at the exact beats in each of its iterations. The colored boxes in the example track the alternating hand patterns. Figure 6.12 shows that even though the visual and haptic functions change, the music stream remains essentially the same and unchanging, therefore solidifying music's role as an intermediary stream that simultaneously checks its sync point relations with gameplay and visual accentuation. The M[P]-[I^2] sync type is *preserved* by the tri-sync line, thus strengthening the presence of metric organization in the closing.

The conformity of Figure 6.12 can also be reflected at the R-type level, as shown in Figure 6.13. As previously discussed, tri-sync often imposes metric sync and de-emphasizes more [R]-type accents like character motion. However, non-metric relations of total sync can still meet trilateral sync, such as the "bye, bye, bye" line in Figure 6.13's m. 41. Here, the eighth notes (music), shot changes (visual), and tap notes (touch-based) fully mickey-mouse each other to emphasize the heightened drama of lyric and dance. This tripartite mickey-mousing pushes

players to hold an active role and physically move between sync points, from figurative point A to point B.

The beatmap's higher degrees of interactivity naturally alters some sync point functions at the tri-sync level. Figure 6.14 exemplifies this with two sync collections for "Close to Gray" that lead to the pre-chorus. It also shows the simplified final product for all the processes featured in Figures 6.9–6.13. The sync train illustrates, without notation, how the M[D]-[I^2] sync collection (Sp.1, 2, and 4) and M[P]-[I^2] reroutes (Sp. 11 and 13) replace prior function labels by virtue of metric hierarchy. Areas that are predominantly [VCO], are still organizational according to vocal fidelity, but the nuanced function changes underscore tri-sync's tendency to highlight metric relations. Tri-sync more often imposes a metric organization based on phrase structure of the music it mirrors. However, as in Figure 6.13, there are moments where tri-sync occurs at the rhythmic level and creates emphatic moments.

With music as a common denominator, sync train construction at the tri-sync level can change functions for sync collections. This mirrors how rhythm gameplay transforms audiences into active participants that create their own synchronous points and contexts. The gameplay interface itself is animation, prompting players towards synchronous precision. In real-time, simultaneously experiencing the gameplay, music, and video can potentially overwhelm the stimuli. Synchronous moments across three steams of combined audio-haptic and audiovisual rhythms help organize and emphasis particular moments such as projecting rhythmic patterns, framing the opening and ending passages, maintaining sonic fidelity, and so forth. Defining synchronous organizational moments is the foremost function of sync point analysis, as applied to rhythm game beatmaps.

6.4: Conclusion

The previous chapters built a framework for conceptualizing how synchronicity dynamizes the filmic experience. Sync points play a marked role in delivering animated sequences, where character actions are deliberately manufactured. Both the sync point analysis and train construction position audiences as receptors of films as finished products, or analytical hypertext. Sync points turn film and gameplay into hypertexts and create interactivity between film and receiver. As this chapter demonstrates, sync point analysis is equally applicable to ludomusicology by providing an objective methodology for studying rhythm game beatmaps.

The figures in this chapter provide a framework for transcribing and notating rhythmic beatmaps for analytical purposes. Transcribed beatmaps are compared against a basic audiovisual transcription, and only moments of triple sync can define or redefine sync point functions, like how the M[P]-[I^2] grouping structures in the opening of "Close to Gray" override and subsume the otherwise R[D]-[ISO] label if it were only music and video. Then, the tri-sync line corrects any out of phase moments created by the audio-haptic rhythms. Transcription and notation can be streamlined and circumvented through use of the sync train, which summarizes the frequency and grouping of sync point types without the need for a music score. Analyzing tri-sync in "Close to Gray," illustrates how the addition of an additional rhythmic stream can lead to alterations in rates of occurrence and function labels. Because rhythm games inherently promote repeated gameplay for mastery, sync point functions help define a beatmap's established "form," or patterning. Constructing tri-sync transcriptions lays the groundwork for analyzing other rhythm game beatmaps of and beyond the *Colorful Stage!* franchise, pushing the utility of sync point analysis beyond the filmic scope.

Chapter 7

Conclusion

My spiraled *sync point analysis framework* and its culmination as *sync trains* aspire to provide an angle of insight into the rhythmic underpinnings of the *esthesic* (reception-oriented) filmic experiences and the *poietic* (production-oriented) stylistic choices of audiovisual rhythm in films and beyond films. Over the course of six chapters, my framework is developed from two angles. First, is rigorous definition of synchronous and non-synchronous points through the n[x]-[y] label. Second, is the establishment of function and its limiters and principles through [AVC] complexes, the five principles of metric sync, the sync train hierarchy, and the trilateral sync and pivot models. My concepts were developed from this dissertation's case studies and refined by analysis. The resulting reverse-engineering of the filmic process not only solidifies my framework but also exemplifies practical application of the apparatus.

Part I, "Taxonomical Toolkit," served as a trilogy of chapters dedicated to defining each element of the n[x]-[y] formula. The technical label's utility is in immediately conveying *what* is being synchronized in the music and *how* it is synchronized through visual accents. Defining *n* as rhythmic or (hyper)metric determines the periodicity and rate of occurrence. The [x] container determines if a [musical accent] is sonic [S], phrasal [P], or durational [D]. The [y] container determines if a [visual accent] is iconic [ICO], isomorphic [ISO], iconic-isomorphic [I^2], or vococentric [VCO]. If the functional sync point occurs at moments of formal closure, it can become an audiovisual cadence [AVC] complex where visual elements reinforce formal logic.

Part II, "Advanced Applications," takes the n[x]-[y] functions to the next level and organizes them into sync trains. These trains are streams of sync points that mark predominant functions, identify formal hierarchies, and dynamize the otherwise static sync points by

insinuating *fabricated motion* between the points. Sync trains describe what happens *between* sync points and contextualize the functions and actions of audiovisual materials. The trains also provide a graphic summary of the filmic process, which not only facilitates analysis and communication, but also serves as a map for reverse-engineering the rhythmic audiovisual elements of a song or scene. In the spirit of sync point analysis, sync trains give large-scale data on sync point frequency, prevalent sync collections, rates of periodicity/non-periodicity, and the level of musical or visual involvement in delivering a scene. The audiovisual involvement is then broken down into three generalizations: totality, visuality, and aurality—each indicative of contextual spaces of fluctuating musical [x] and visual [y] markedness.

The focal case study of Part II, "Close to Gray" (*Colorful Stage!*, 2021), exemplifies the analytical potential of sync trains. The train condensed all the sync points of "Close to Gray" into a graph to illustrate four main types of sync functions, a predominance of non-periodicity and R-sync (wavy lines), and a predominance of shot changes (filled nodes) used in the Nightcord at 25:00 music video. Totality, or equal emphasis of musical and visual features, connects the sync points and establishes an economical ecosystem of sync types whereby any *reroute*, or secondary emphasis of sync types, uses one of the four sync types found in the train. The Nightcord version is then compared to the *Vocaloid* version by Surii—one that features more sync collections, offers no choregraphed performance, and de-emphasizes totality in favor of highlighting music over image, or vice versa, at different times.

Just like the Chapter 2 comparison of Disney's and Warner Brothers' adaptations of *Carnival of the Animals*, the comparison between Nightcord's and Surii's versions of "Close to Gray" exemplify how sync point analysis and trains can summarize stylistic and narrative differences through technical means. The sync train and principles of metric sync were applied to

Chapters 5 and 6's spiraled case study of "Close to Gray" to show that the framework, in what is presently its most advanced state, underscores three factors:

1. that sync analysis of emergent rhythms can iron out potential contradictions among differing sync types.
2. that tri-sync fluctuates between contradictory and non-contradictory moments to weave an organic whole.
3. that metric-, phrase-oriented, and out-of-phase beatmap patterns can yield the same rhythmic profile regardless of different physical/tactile or animated choreographies.

The development of these three factors illustrates the sync point analysis' utility, even when extended to multimedia beyond films.

Within the scope of this dissertation, my sync point framework primarily contributes a method for analyzing animated songs that are separable from (and marketable for) their feature film or created as self-contained music videos. Case studies of Disney's sing-alongs from films like *Frozen* (2013), *Hercules* (1997), *Mulan* (1998), *Moana* (2016) applies sync point analysis to feature songs that emphasize an ideal or emotion away from the unfolding narrative. Case studies of "Cinema" (2021) or "Close to Gray" (2021) from the *Colorful Stage!* series (2020) are music videos with their own self-contained narratives, but sync point analysis demonstrates how both the sing-alongs and music videos share the same stylistic uses of functional sync points.

For my case studies, the music is preexisting, like that of the "Finale" from *Carnival of the Animals* or commissioned for animated adaptations in the case of "Close to Gray." This slightly prioritizes music as a precedent and preexisting source to which the images are timed to. That is why music [x] serves both as indicators of *what* images are synchronized to as well as the

seams connecting tri-sync. However, assuming that music is the sole precedent for synchronicity overlooks the production-oriented nature of films. As discussed by Disney animators Frank Thomas and Ollie Johnston, matching animation to music might also require changes to music written specifically for the films.[187] The changes might be unknown to the filmgoer and includes factors such as the addition/subtraction of musical beats and measures, the use of "dummy" or temp tracks, music that is not prescored, or music that is changed last-minute.[188] Therefore, my framework seeks to accommodate even those changes by placing music and image in a reciprocal relationship, and the use of graphic representations of beat notation and sync train nodes to emphasize image relations to a musical beat if meticulous attention to the musical surface is not invoked by the visuals [y].

The flexibility of my framework opens possibilities of analyzing beyond the standalone Disney numbers and animated music videos. As demonstrated by the case studies of *Steamboat Willie* (1928) and *Alexander Nevsky* (1938), sync point analysis can apply to "unsung" or "voiceless" scenes that focuses on different timbres (*Steamboat*) or orchestral scores that set narrative atmosphere (*Nevsky*) in live action films of which Eisentsein's original montage techniques were conceived for. Therefore, future avenues for research includes plans to develop my analytical system into one that can also address other areas of the animated feature film that are not standalone songs, feature less use of sync points, or are cinematic versions of a standalone song's leitmotif. Future research can also test the framework on other forms of media, such as the live action musical, karaoke video with kinetic typography, or even synchronous relationships between theme park music and parkgoers.

---

[187] Frank Thomas and Ollie Johnston, *the Illusion of Life: Disney Animation* (New York: Hyperion, 1995), 287–297.
[188] Ibid.

In the end, my sync point analytical framework is a core set of parameters and empty containers. It can adapt to different animated multimedia and diverse repertoire, while being a flexible and learnable apparatus. The framework's intention is not to oversimplify the esthesic and poietic underpinnings of film, but to understand, frame, and describe the focal elements that make a scene work. Through this understanding, animated musicals, music videos, and rhythm gameplay can be deeply appreciated for their technicalities and intricacies in delivering an emergent rhythmic narrative and immersive, synesthetic experience.

Naturally, words and functional labels can only scratch the surface of the filmic experience, let alone describe all the synchronous possibilities known and unknown to filmic space. Yet, I believe that sync point analysis that is interdisciplinary and informed can help us understand what, why, and how animated moving pictures are worth a thousand (or more) words. After all, synchronous relations drive, enliven, and organize our reception of animated audiovisuality. The eternal surprise is that sync points in cinema can be awaiting us at any moment—where you are, beyond open doors, far under the sea, or even close to gray.

Part III: Appendix

Figures

Chapter 2 Figures

Figure 2.1: Visual and audio (musical) synchronicity in opening of *Flamingos* (2000). Reduction by the author.

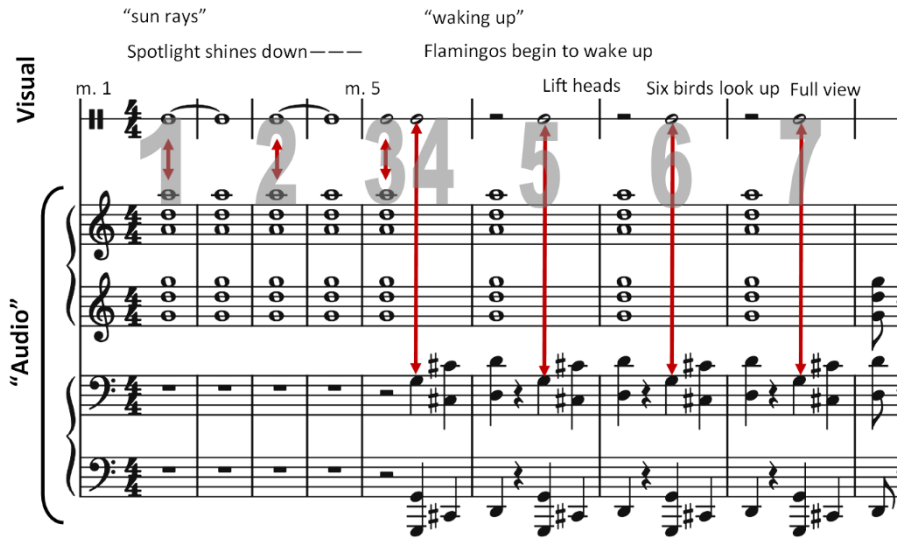
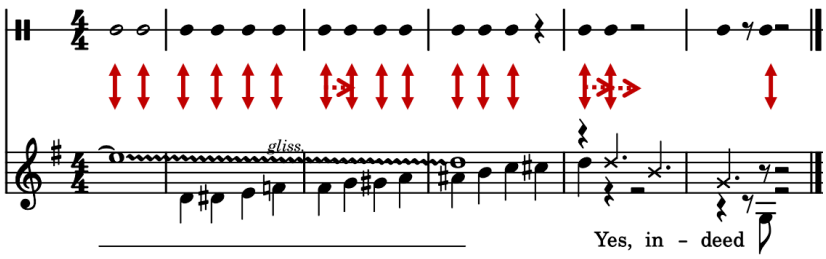


Figure 2.2: Sync points in the ending of "Zero to Hero" from *Hercules* (1997). Transcription by the author.[189]



[189] Watch "Zero to Hero" here: https://youtu.be/yOL-EJZjmp0?si=ML1_zcJfz7u1yf04. Excerpt begins at 2:04.

Figure 2.3: Sound effect (tree branch collision) represented in piano reduction from *Flamingos* (2000). Reduction by the author.

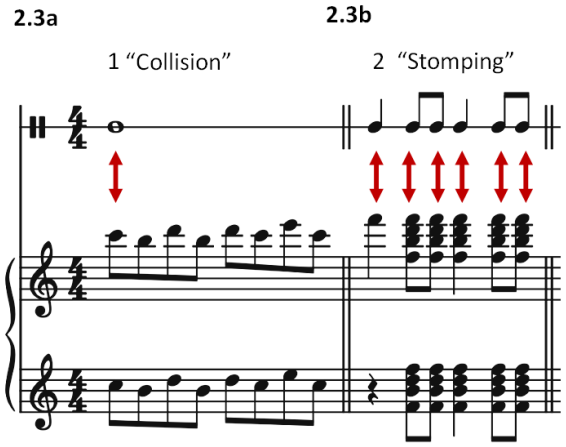


Figure 2.4: The Sync Point Analysis plugin formula

n[x]-[y] = Sync rate x [Music accent type]-[Visual accent technique]

n = sync rate = rhythmic rate of occurrence
[x] = [music accent rate]
[y] = [visual accent type]

Figure 2.5: An R[S] sync point. How a single sonic event is visually emphasized. “Finale,” mm. 57–59.

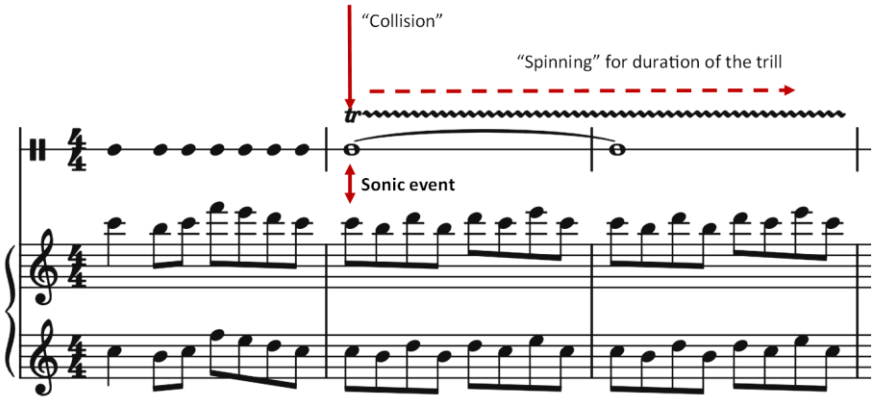


Figure 2.6. An R[P] relation in "Love is and Open Door" from *Frozen* (2013), showing how visuals emphasize an entire musical phrase. Transcription by the author.[190]

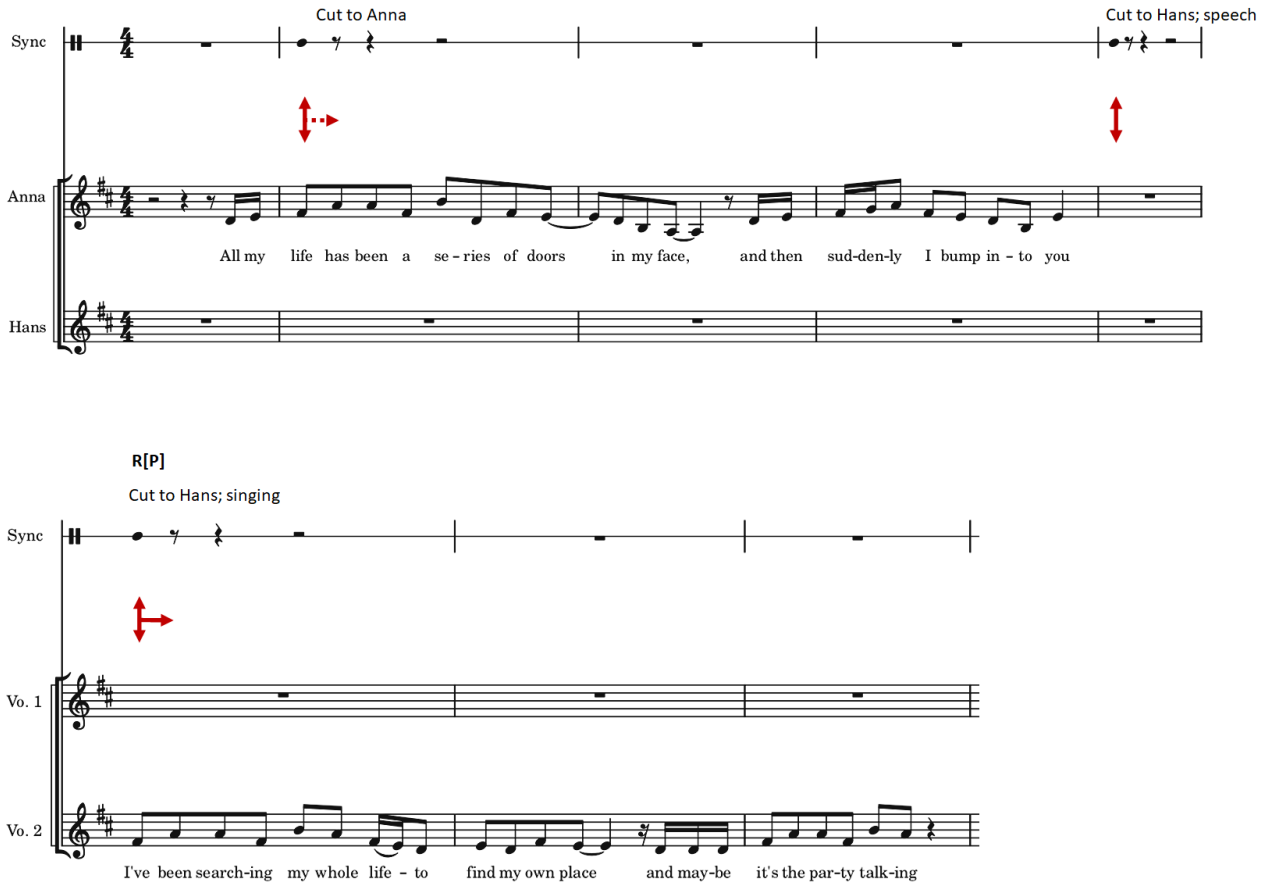
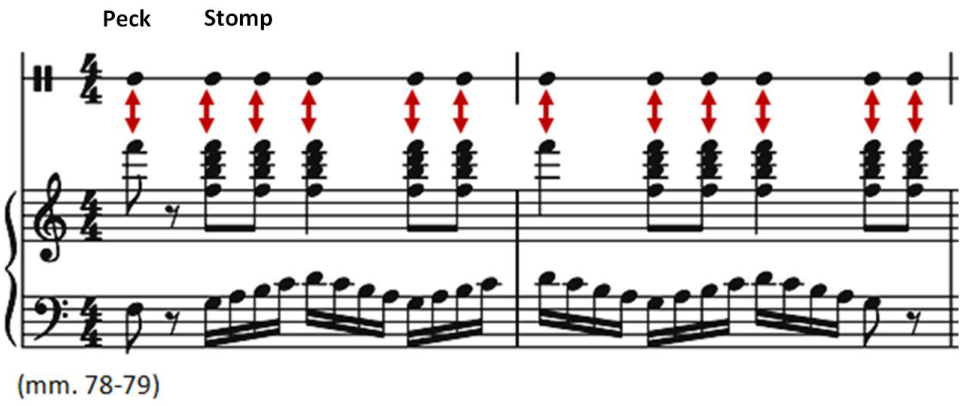


Figure 2.7: An R[D] relation. How visuals emphasize duration over pitch in *Flamingos* (2000).



[190] Watch "Love is an Open Door" here: https://youtu.be/kQDw88hEr2c?si=V-ixJYjIKXOb64Sf. Excerpt begins at 0:06.

Figure 2.8: Summary of visual accent type indicators and categories.

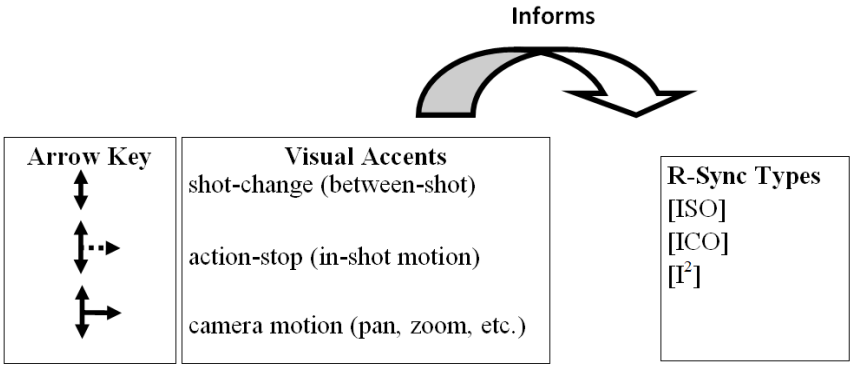
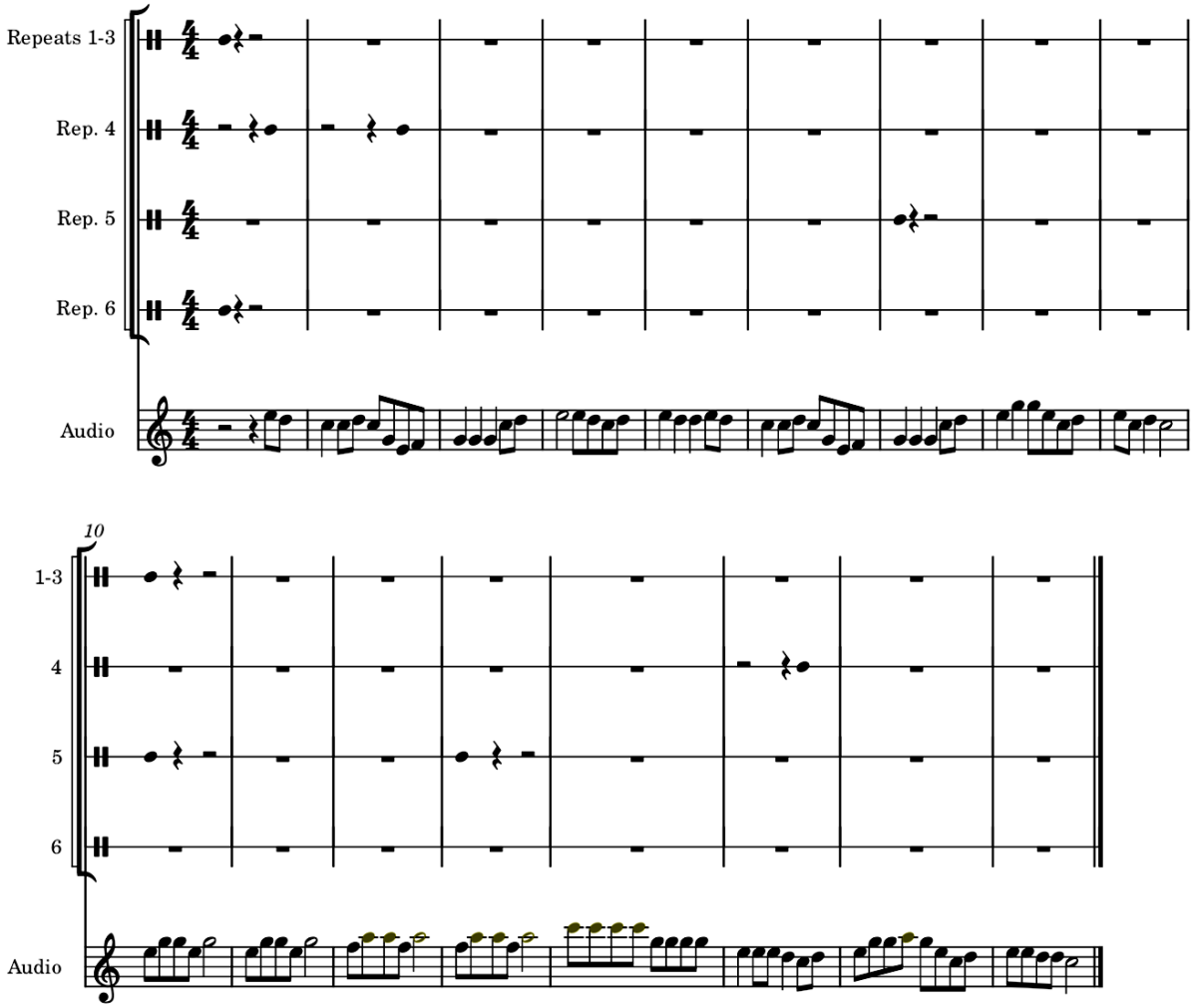
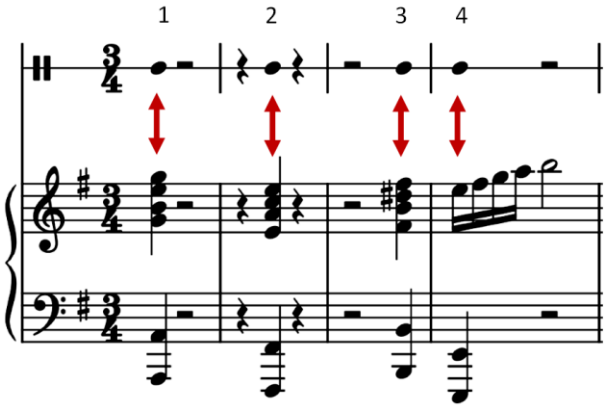


Figure 2.9: Slow rate of shot-changes in *Steamboat Willie* (1928). Transcription of "Turkey in the Straw" by the author.[191]



[191] Watch *Steamboat Willie* here: https://youtu.be/BBgghnQF6E4?si=AJdVQ77eQso8C0h9. Excerpt begins 4:22.

Figure 2.10: Shot changes that mirror musical, harmonic and cadential motion from "Hokori to Ogori" from *Shoujo Kageki Revue Starlight* (2018). Transcription by the author.



Events:
1. Shot change; Maya's hand reaches up
2. Shot change; close up behind Maya
3. Shot change; extreme close up of eyes
4. Shot change; full shot of Maya; smoke towers ignite behind Maya

Figure 2.11: Camera motion as visual accents in "Honor to Us All" from *Mulan* (1998). Transcription by the author.

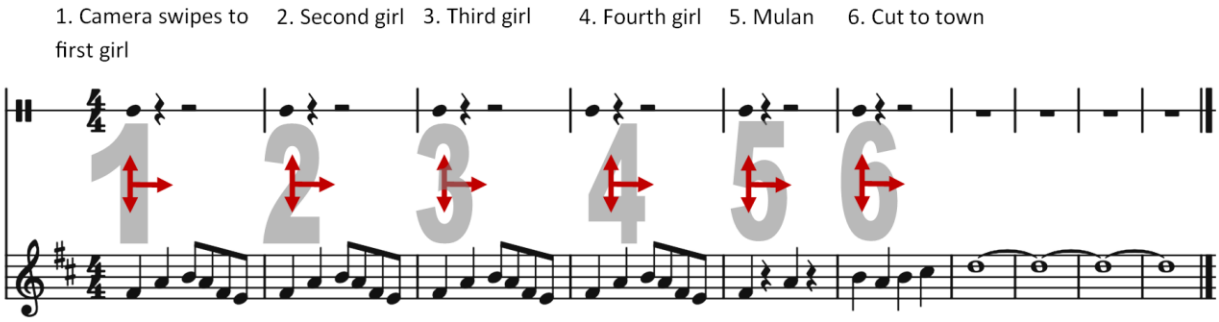


Figure 2.12: Combinations for the sync point formula.

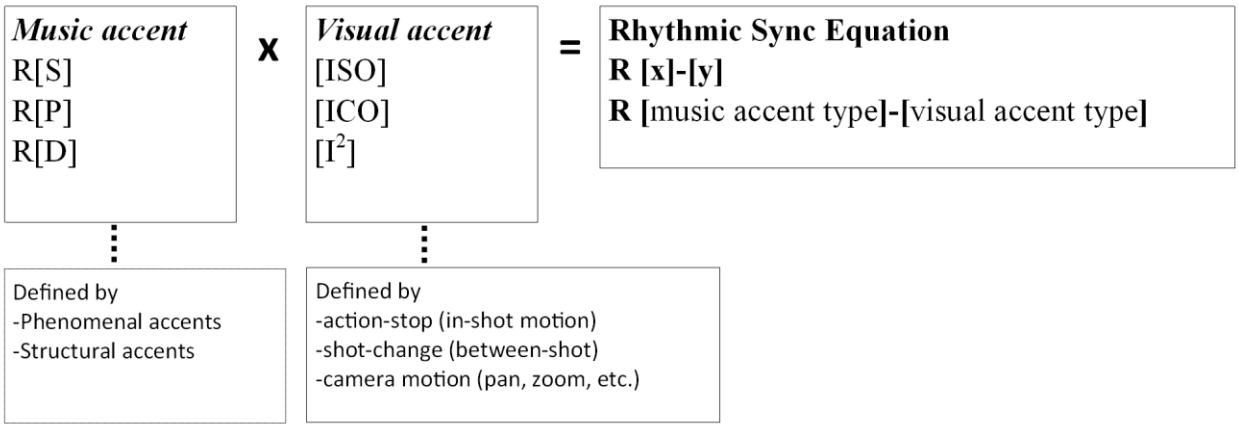
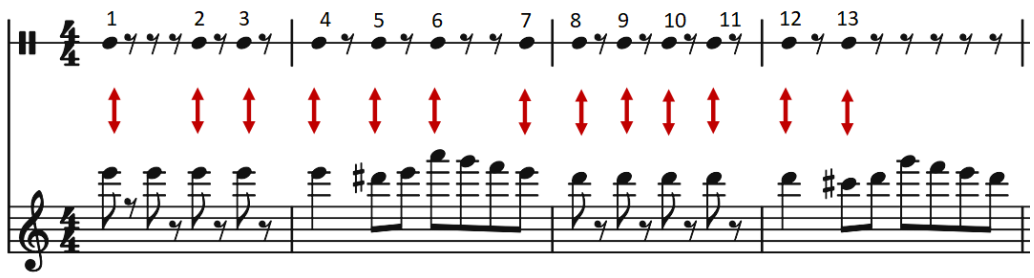
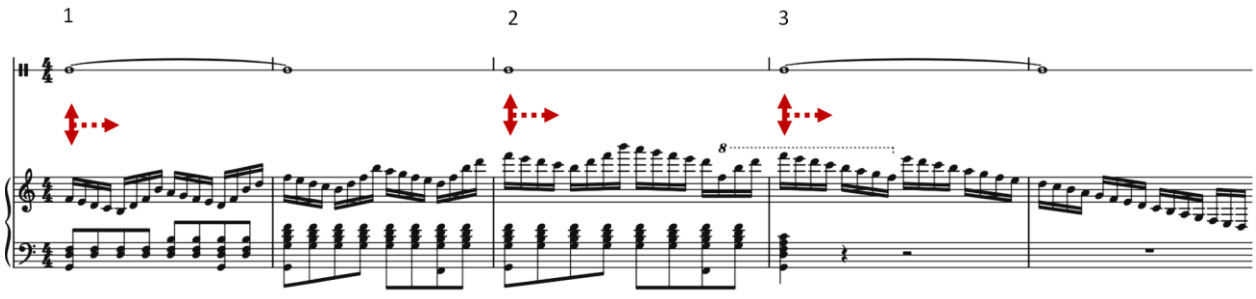


Figure 2.13: Examples of R[D]-[ISO] sync from *Bugs and Daffy's Carnival of the Animals* (1976), mm. 45–48.



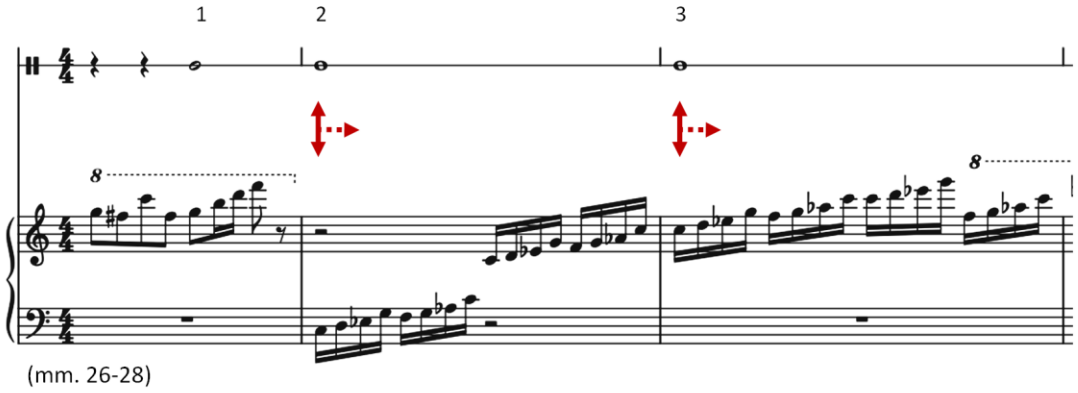
Events:
1. Violin superimposed over piano keys
2. Keyboard backdrop disappears; violin scroll remains as a still image
3. Shot change. Three violins appear.
4. Shot change. Medium close up of violin.
5. Shot change. Medium close up to Violin waist.
6. Shot change. Extreme close up of violin strings.
7. Shot change. Extreme close up of wood. Violin is unrecognizable.
8. Shot change. Close up to French horn bell
9. Shot change. Extreme close up of French horn valves and leadpipe.
10. Shot change. Extreme close up of valves.
11. Shot change. Another close-up of lead pipe.
12. Shot change. Full shot of French horn.
13. Shot change. Focus shifts to trombones.

Figure 2.14: Examples of R[S]-[ISO] sync from *Bugs and Daffy's Carnival of the Animals* (1976), mm. 39–42.



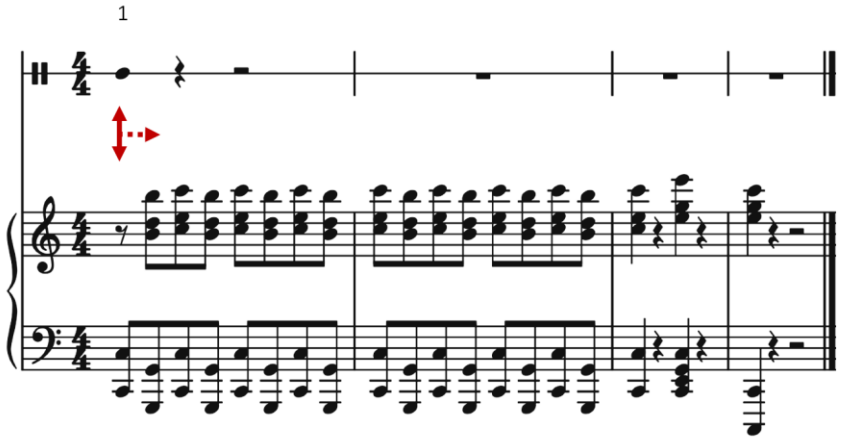
Events:
1. Single line drawn on screen, spun leftward to make a vertical line.
2. Black keys appear one by one.
3. Colors fill out the keys.

Figure 2.15: In-shot, action-stop accents in an R[S]-[ICO] from *Flamingos* (2000), mm. 26–28.



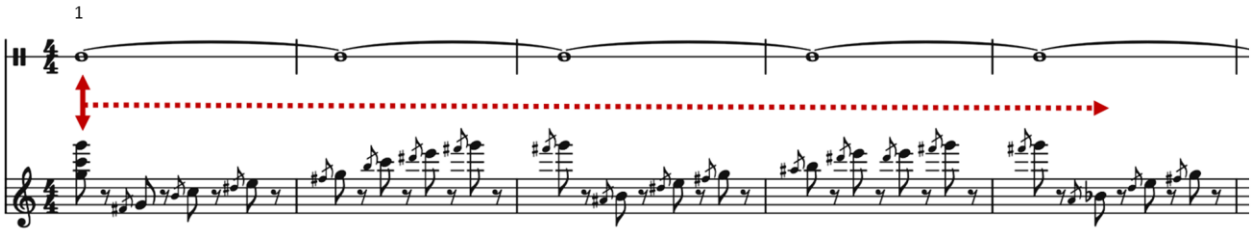
Events:
1. Camera scrolls from the water-skating flamingos to the yoyo flamingo
2. Camera stops at yoyo flamingo; screen shakes; flamingo plays with yoyo
3. Flamingo continues to play yoyo; lifts head and smiles

Figure 2.16: R[P]-[ICO] in *Bugs and Daffy's Carnival of the Animals* (1976), mm. 88–91.



Events:
1. Cut to Michael Tilson Thomas conducting the final cadence; the camera stays fixed on the conductor until the end.

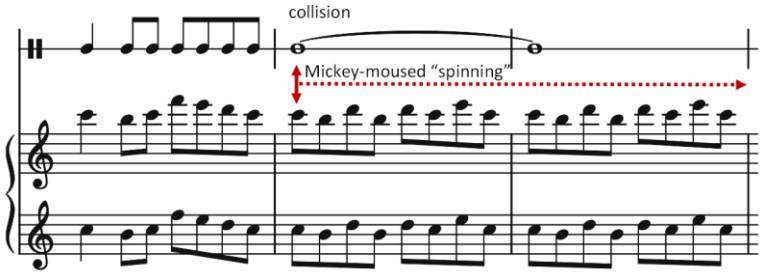
Figure 2.17: R[D]-[ICO] in *Bugs and Daffy's Carnival of the Animals* (1976).



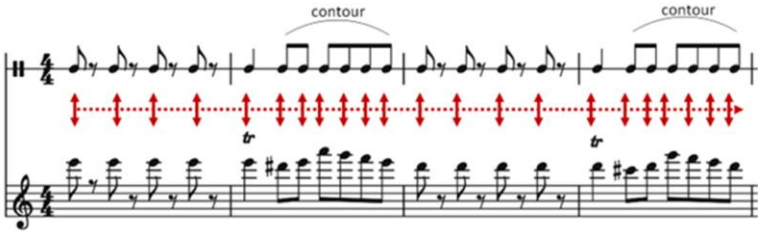
Events:
1. *Initiation* of the large kangaroo jumps. (Tied notes represent the duration of the jumps.)

Figure 2.18. R[x]-[I^2] examples from *Flamingos* (2000).

2.18a



2.18b



2.18c

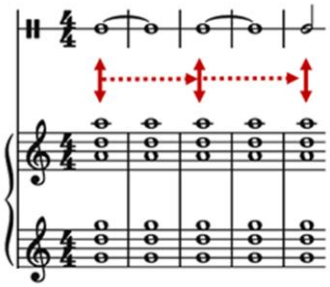
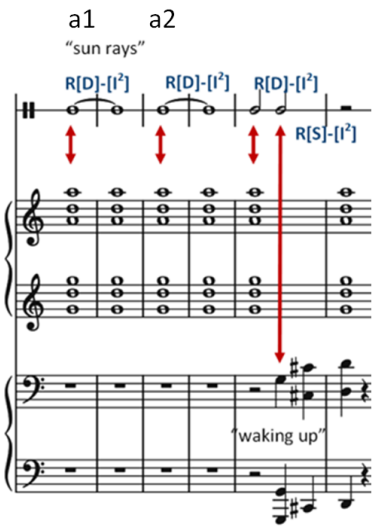
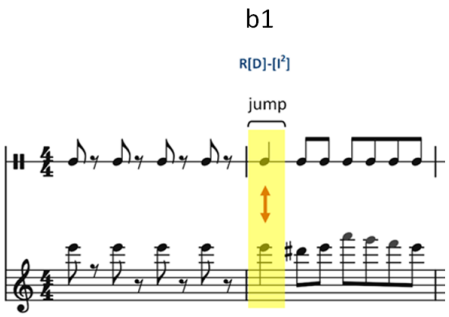


Figure 2.19: Major R[D] instances in *Flamingos* (2000).

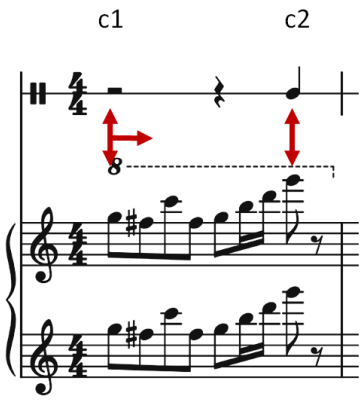
2.19a. (mm. 1–5)



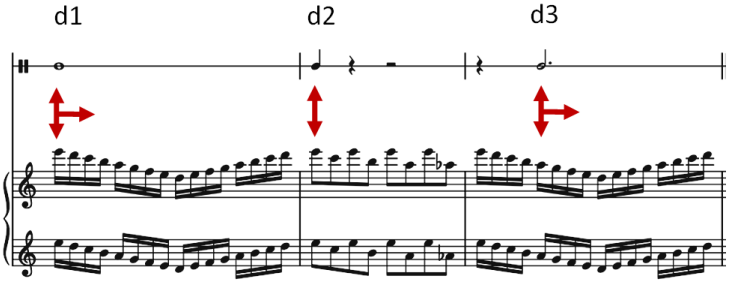
2.19b. (mm. 11–12)



2.19c. (m. 26)



2.19d. (m. 81—83)



2.19e. (mm. 86—87)

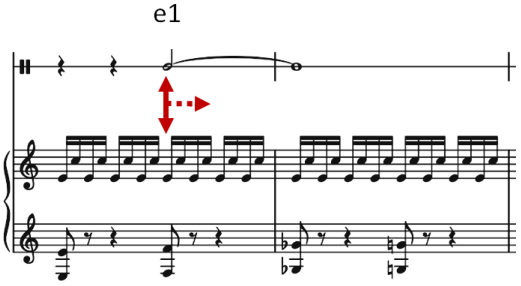
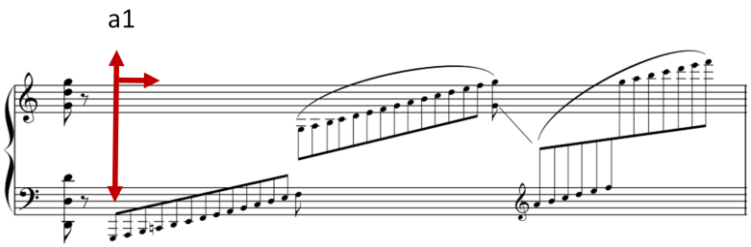
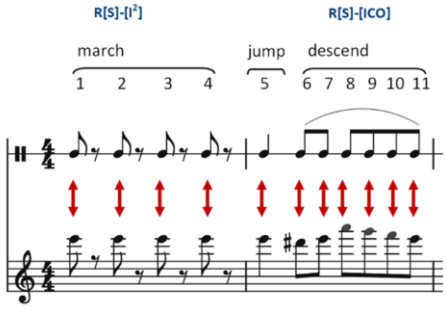


Figure 2.20: Major R[S] instances in *Flamingos* (2000).

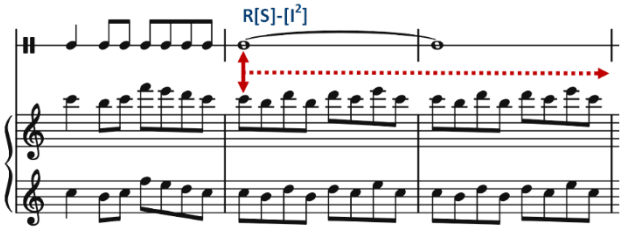
2.20a. (m. 9)



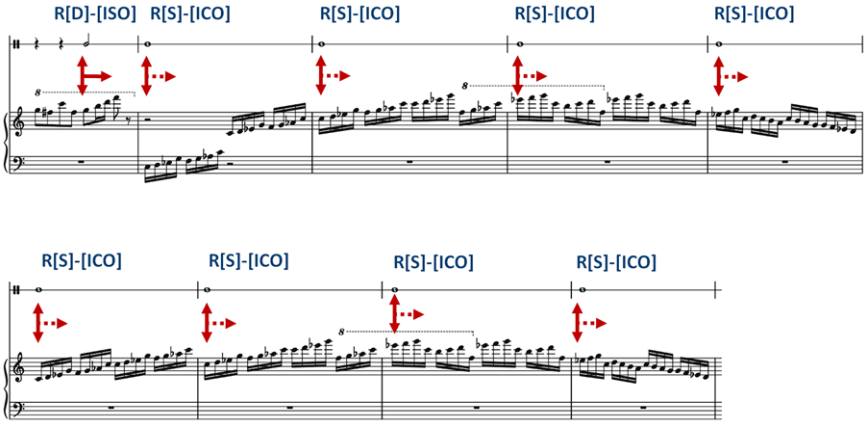
2.20b. (mm. 11–12)



2.20c. (m. 58)



2.20d. (m. 26–34)



2.20e. (m. 39)

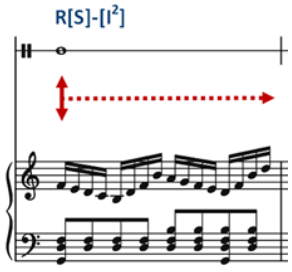
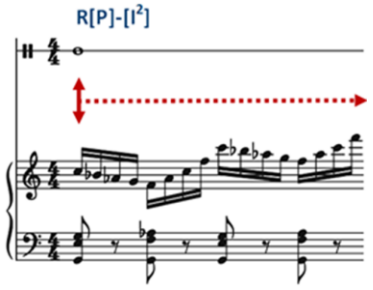
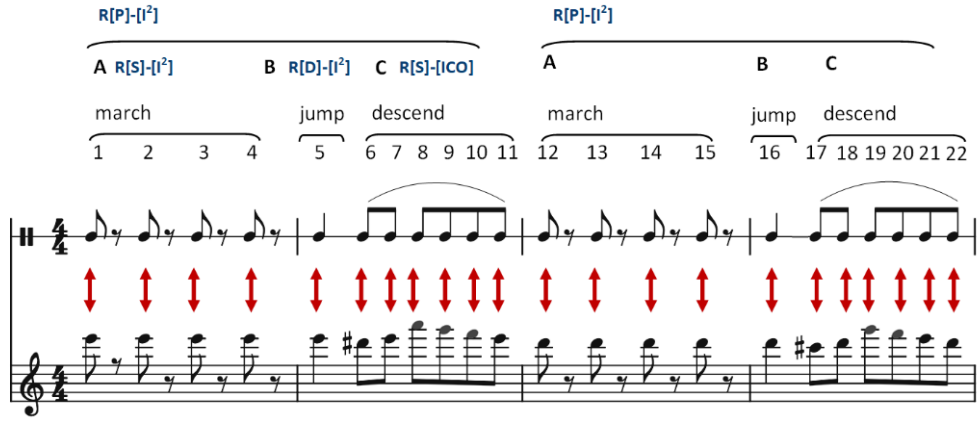


Figure 2.21: Major R[P] instances in *Flamingos* (2000).

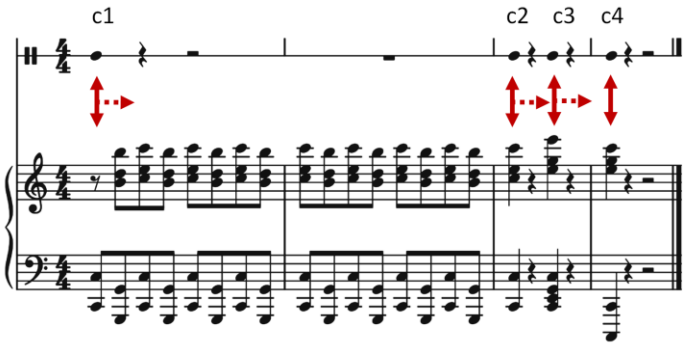
2.21a. (m. 35)



2.21b. (mm. 11–14)



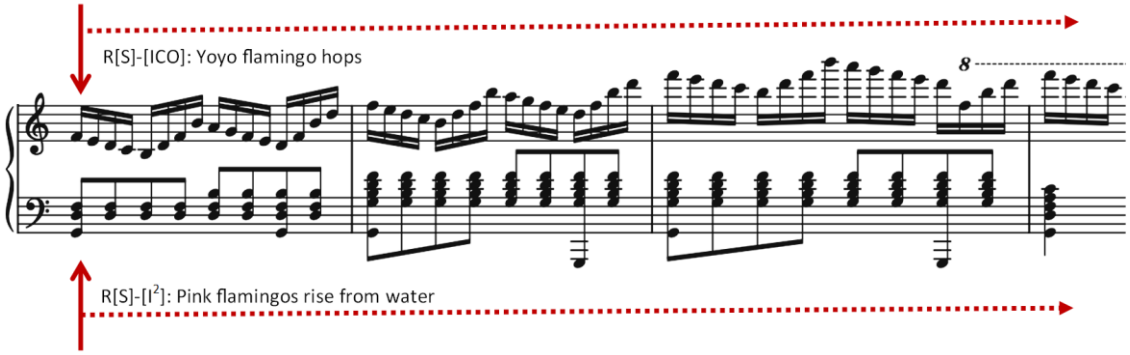
2.21c. (m. 88)



Events:
C1. Yoyo flamingo burst from water; spins yoyos with both wings; spins common flamingos atop head.
C2. Action-stop; pose.
C3. Action-stop; pose.
C4. Action-stop; purple yoyo swings into and blacks out screen.

Figure 2.22: Narrative formed from sync point analysis of *Flamingos* (2000).

2.22a. (mm. 39–41)



2.22b. (mm. 62–66)

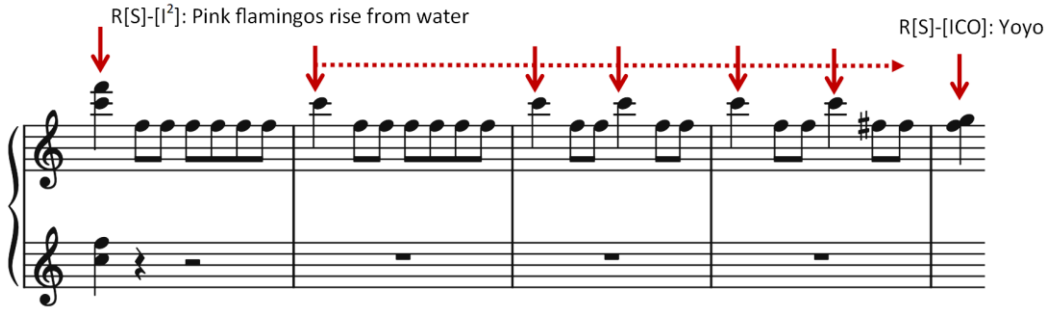
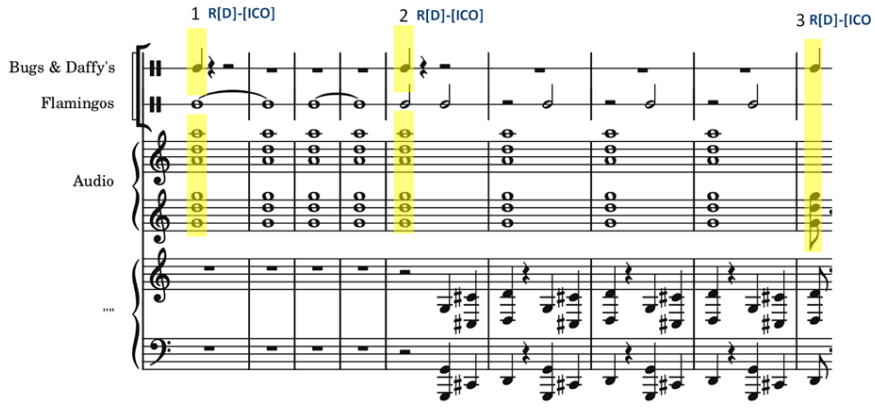


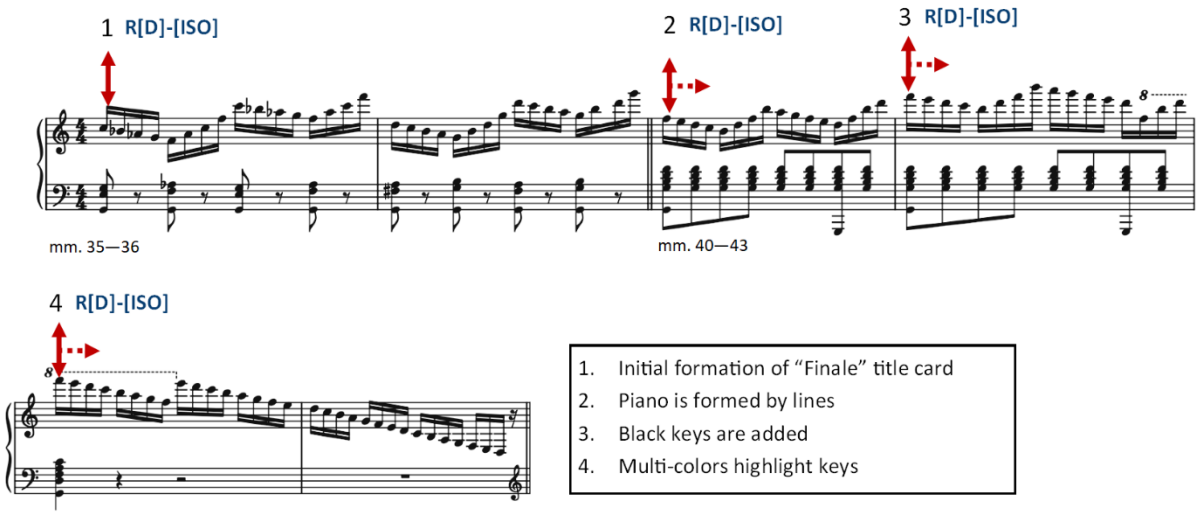
Figure 2.23. R[D] sync types in *Bugs and Daffy's Carnival of the Animals* (1976).

2.23a. Opening (mm.1–9)

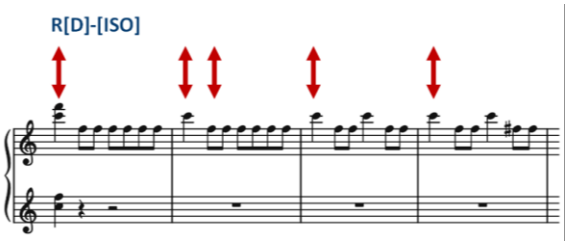


Events:
1. Shot change; close up to Michael Tilson Thomas (side profile)
2. Shot change; wide shot of orchestra
3. Shot change; medium close up of Michael Tilson Thomas (front profile)

2.23b. Title Card Sequence



2.23c. mm. 62–65



2.23d. mm. 13–14

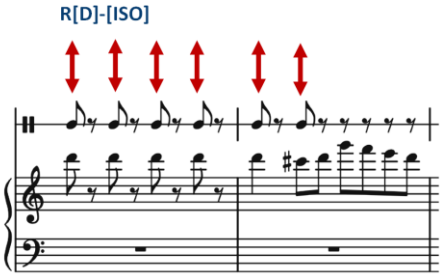


Figure 2.24: The only appearance of Bugs Bunny and Daffy Duck.

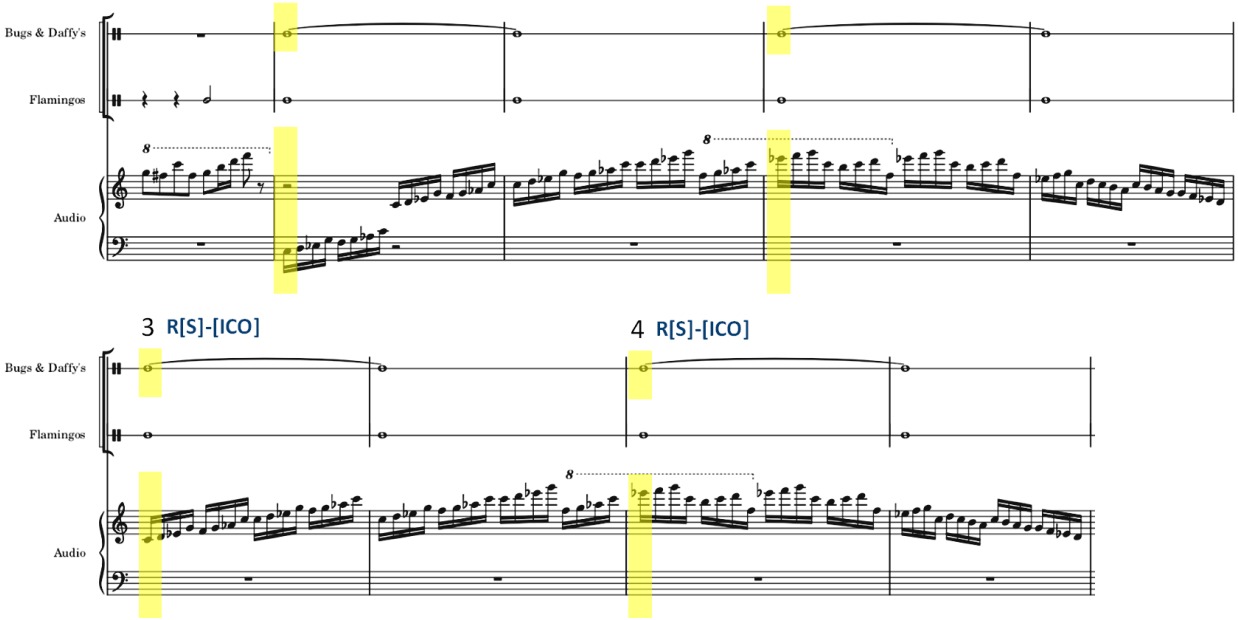


Figure 2.25: Kangaroos with two simultaneous sync types and rhythms from *Bugs and Daffy's Carnival of the Animals* (1976).

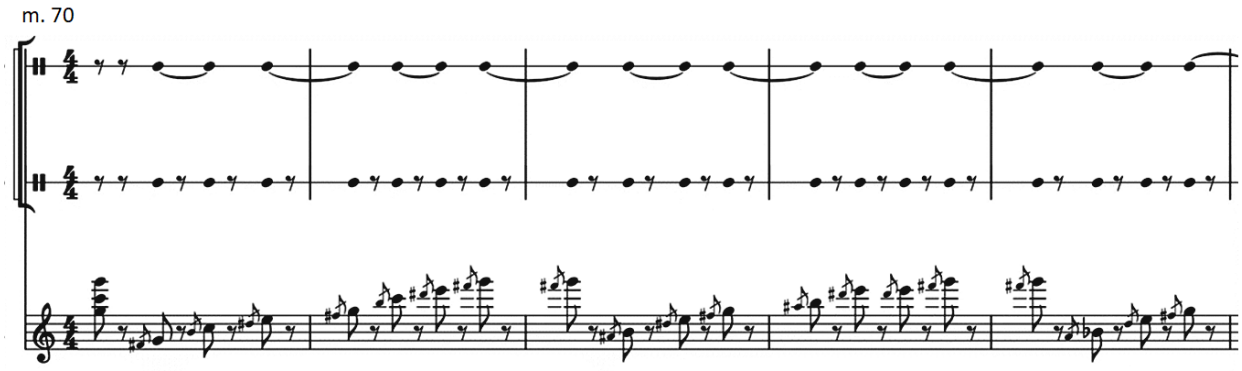
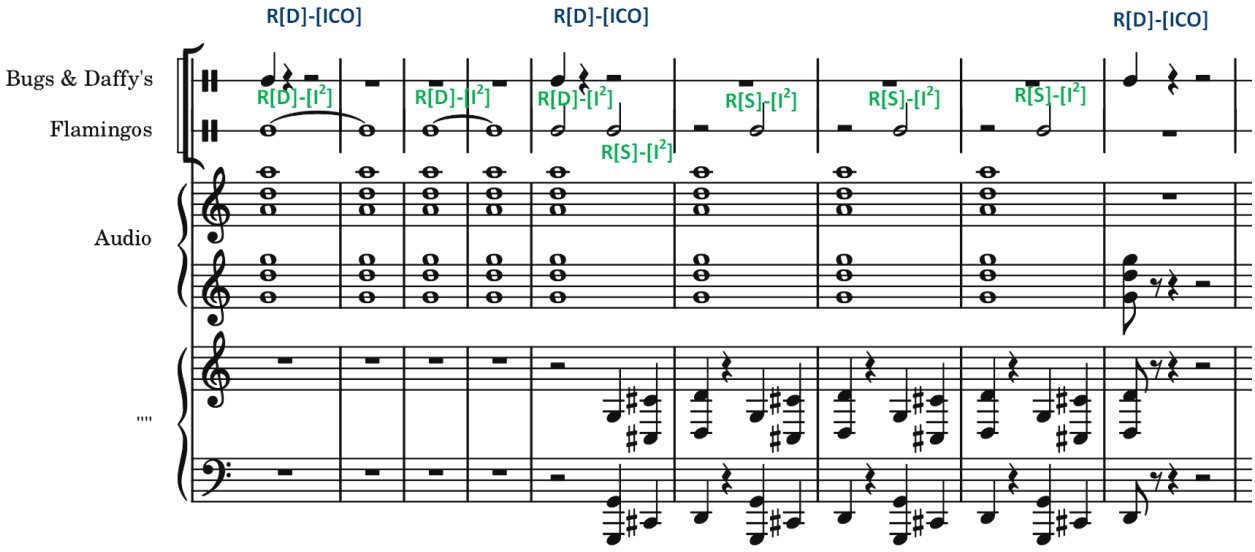
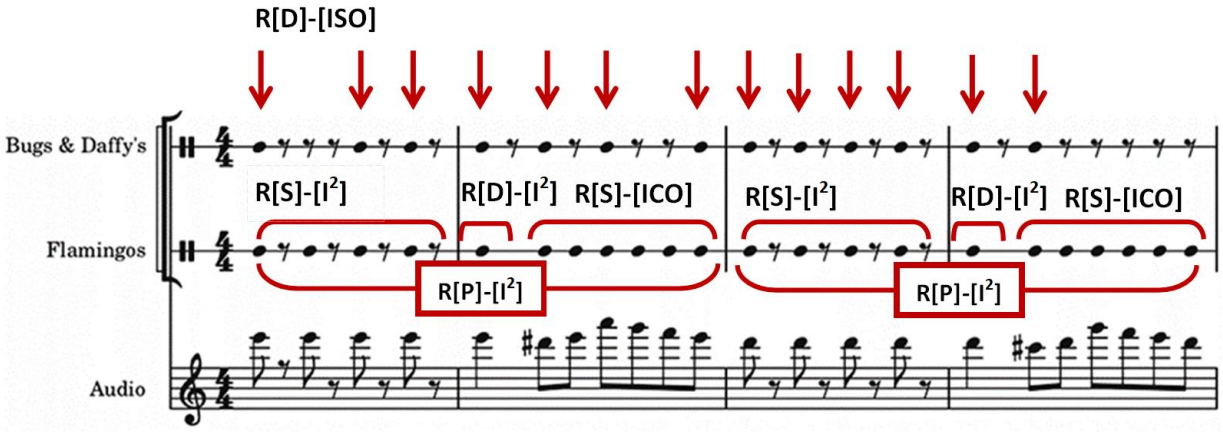


Figure 2.26 : Comparative sync point analyses.

2.26a. Comparison of opening point types.



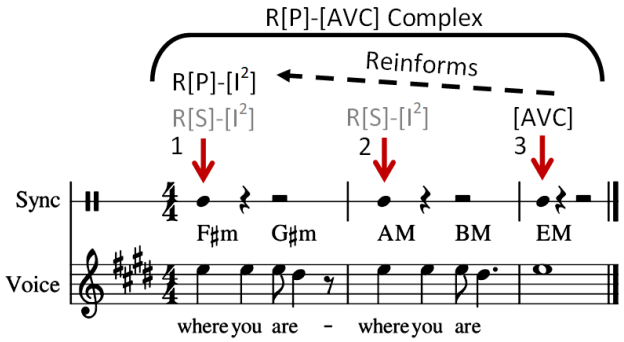
2.26b: Comparison of main theme sync point rhythm.



Chapter 3 Figures

Figure 3.1: Three Examples of [AVC] cadences.[192]

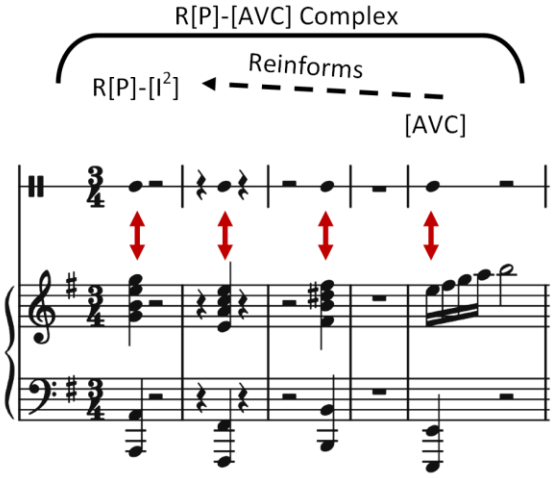
3.1a. "Where You Are," *Moana* (2016)



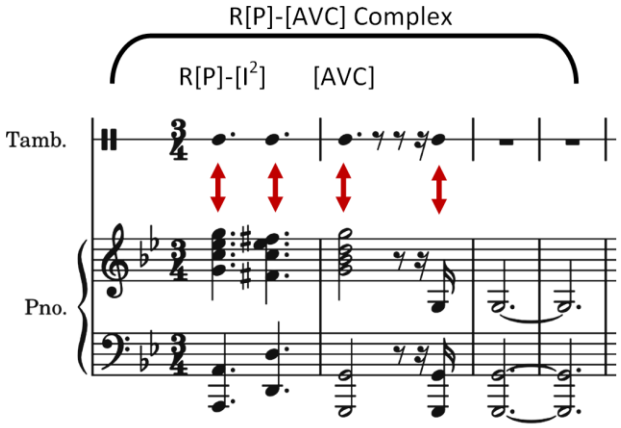
Sync Points

1. Cut to wide shot of Moana, family, and villagers
2. Cut to aerial shot of island
3. Cut to extreme wide shot of island

3.1b. "Hokori to Ogori," *Revue Starlight* (2018)



3.1c. "Hokori to Ogori," *Revue Starlight* (2018)



[192] Watch "Where You Are" here: https://youtu.be/RTWhvp_OD6s?si=Eud1g1K2w7GbzSv9. Excerpt begins at 3:14.

Figure 3.2: Reannotation of [AVC] cadences in the Disney and WB adaptations of "Finale" from *Carnival of the Animals*.

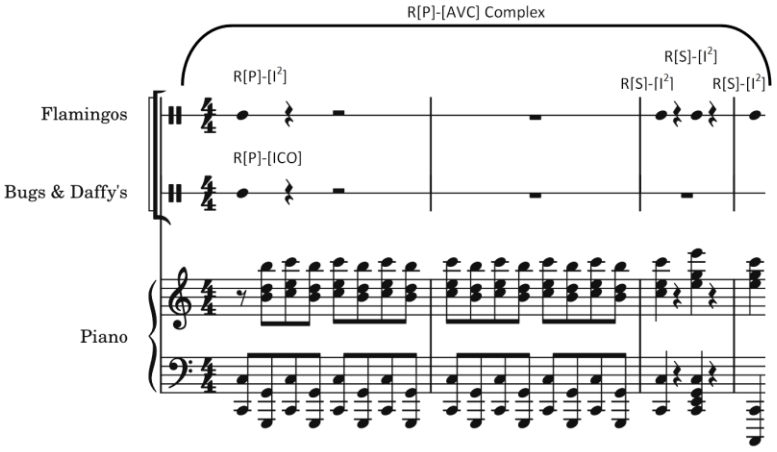


Figure 3.3: Final [AVC] in "Poor Unfortunate Souls" from *The Little Mermaid* (1989).[193]

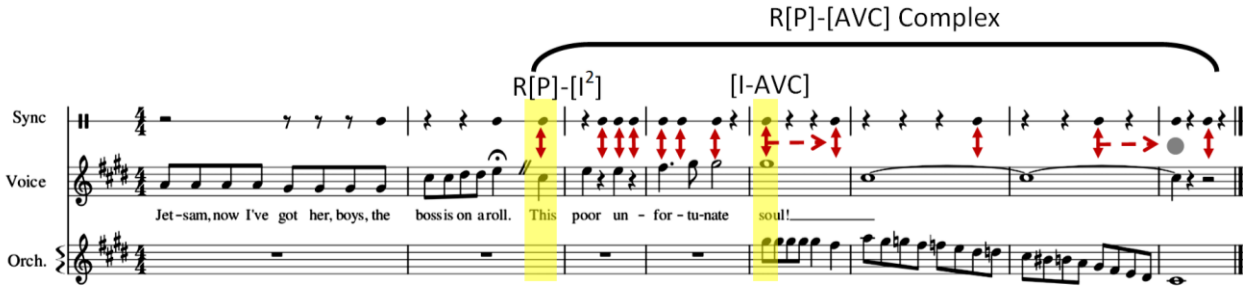
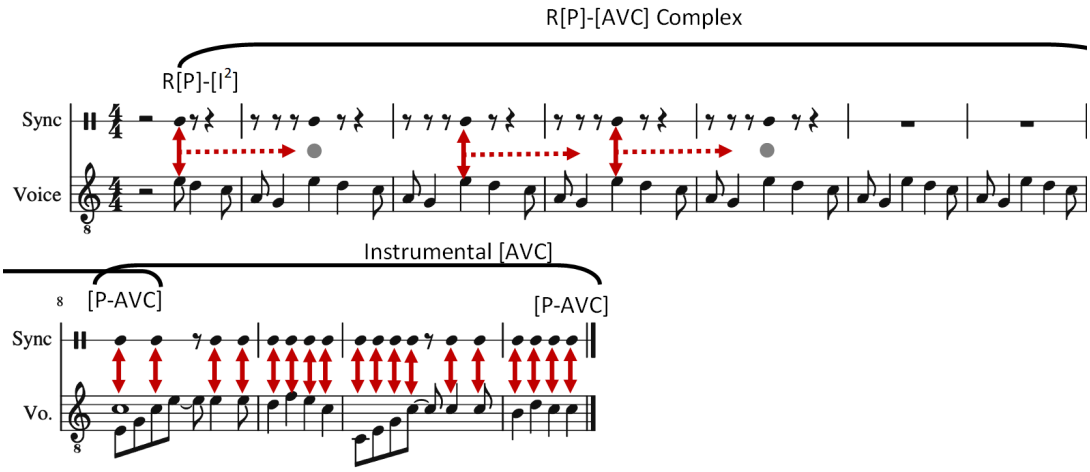


Figure 3.4: Final [AVC] in "Under the Sea" from *The Little Mermaid* (1989).[194]



[193] Watch full video of "Poor Unfortunate Souls" here: https://youtu.be/Gi58pN8W3hY?si=5jpwMKmh2flgzruR. Excerpt begins at 4:11.
[194] Watch full video of "Under the Sea" here: https://youtu.be/GC_mV1IpjWA?si=fgue5Wp2HbEi5RKl. Excerpt begins at 3:01.

Figure 3.5: Final [AVC] in "Honor to Us All" from *Mulan* (1998).

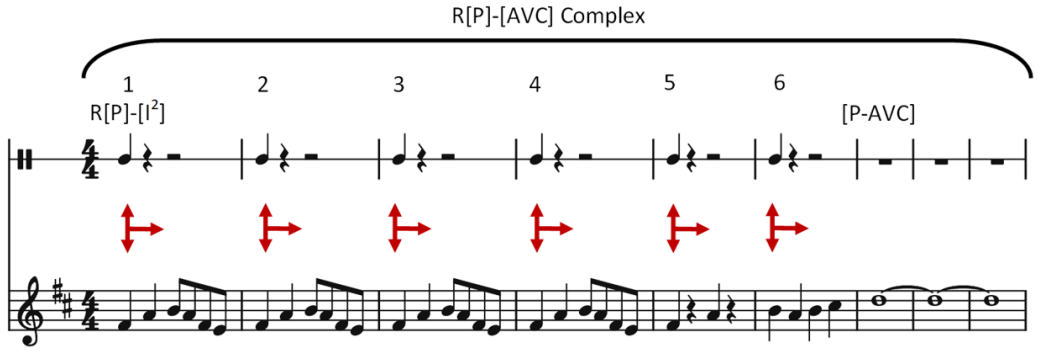


Figure 3.6: Final [AVC] in "Hellfire" from Hunchback of Notre Dame (1996).

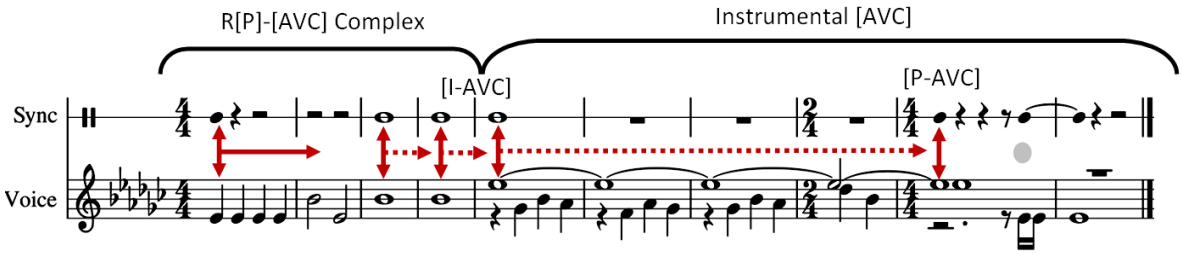
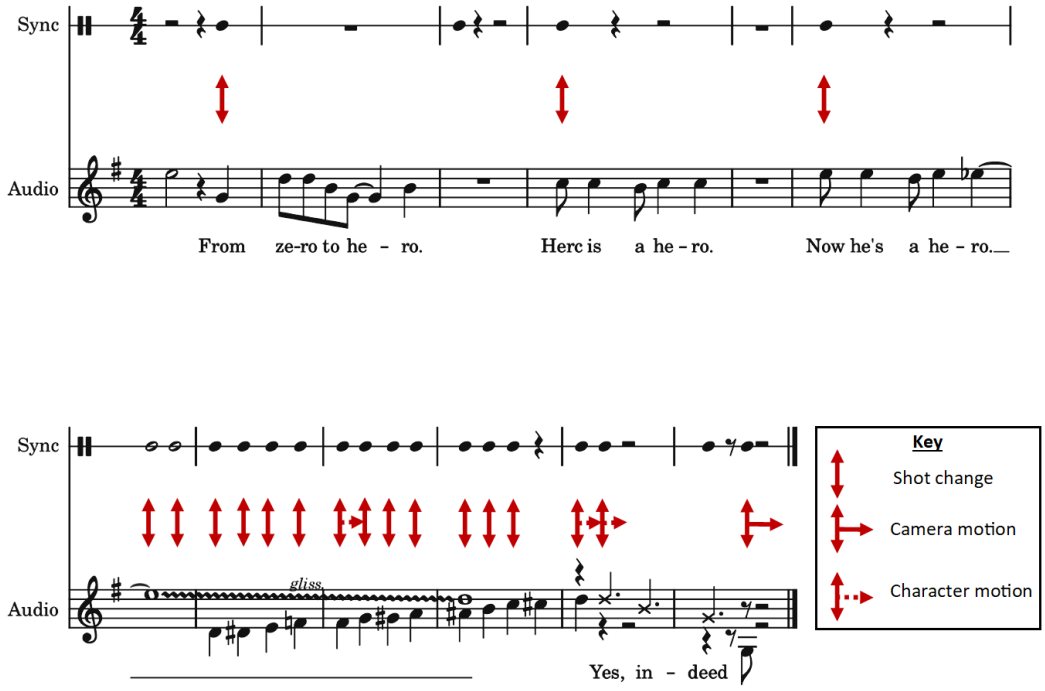
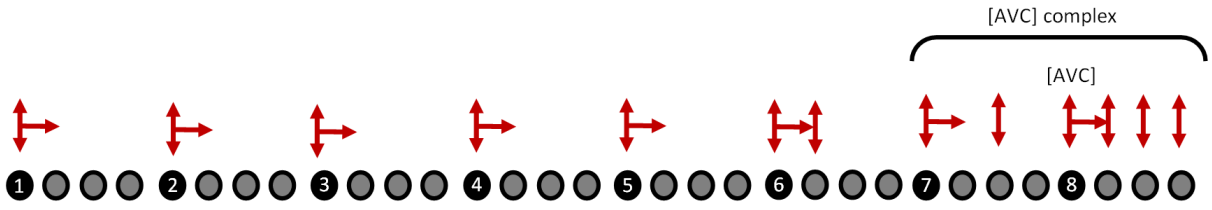


Figure 3.7: Final [AVC] in "Zero to Hero" from *Hercules* (1997).[195]



[195] "Zero to Hero": https://youtu.be/yOL-EJZjmp0?si=SuUAni1ef_Y8R0Lx. Excerpt begins at 2:02.

Figure 3.8: Identifying [AVC] in Scoreless Examples of "Cinema" (2021) from *Hatsune Miku: Colorful Stage!* (2020), also known as *Project Sekai: Colorful Stage!* in Japan.[196]



Events:
1. Shot change to An
2. Zoom out to Kohane
3. Zoom out to Akito
4. Zoom out to Touya
5. Zoom out to Kaito
6. Swipe transition to Akito
7. Shot change to theater seats

Figure 3.9: R[S]-[VCO] in "Zero to Hero" from *Hercules* (1997).[197]

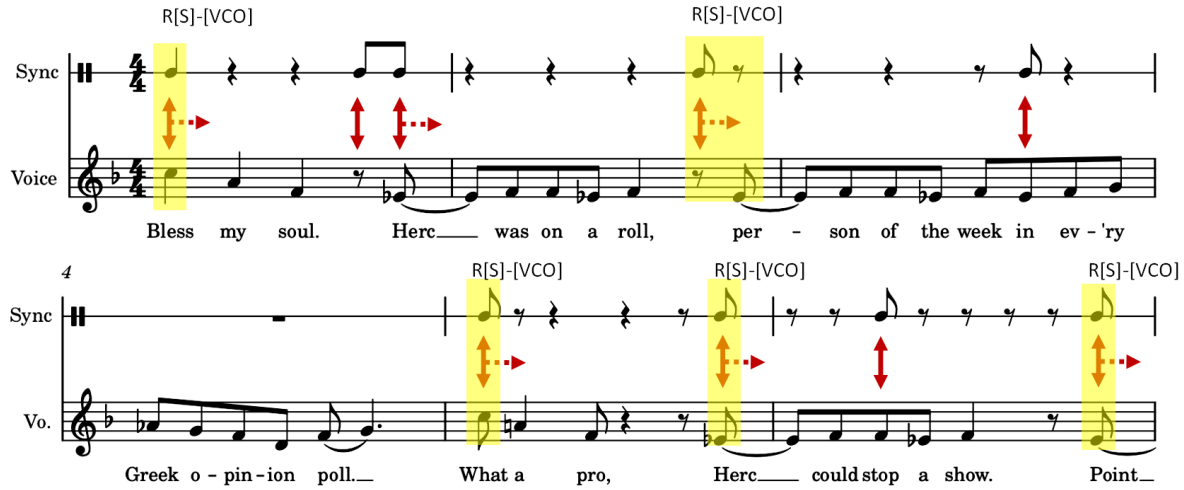
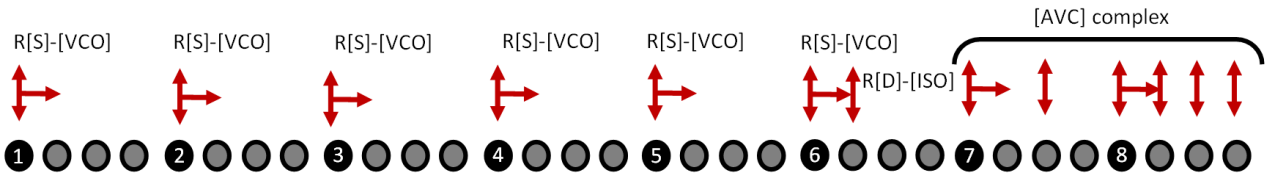


Figure 3.10: [VCO] in "Cinema" from *Hatsune Miku: Colorful Stage!* music video.



[196] Watch "Cinema" here: https://youtu.be/8m1cUgLepmY?si=Pgkg6wg69kS9nKWi. Excerpt begins 1:19.
[197] Watch "Zero to Hero" here: https://youtu.be/yOL-EJZjmp0?si=ML1_zcJfz7u1yf04.

Figure 3.11: [VCO] in “Love is an Open Door” from *Frozen* (2013).[198]

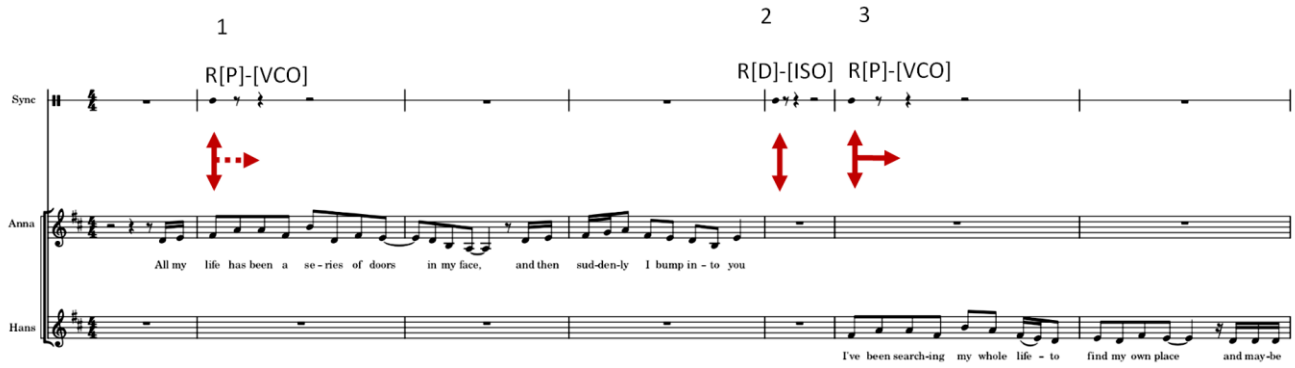
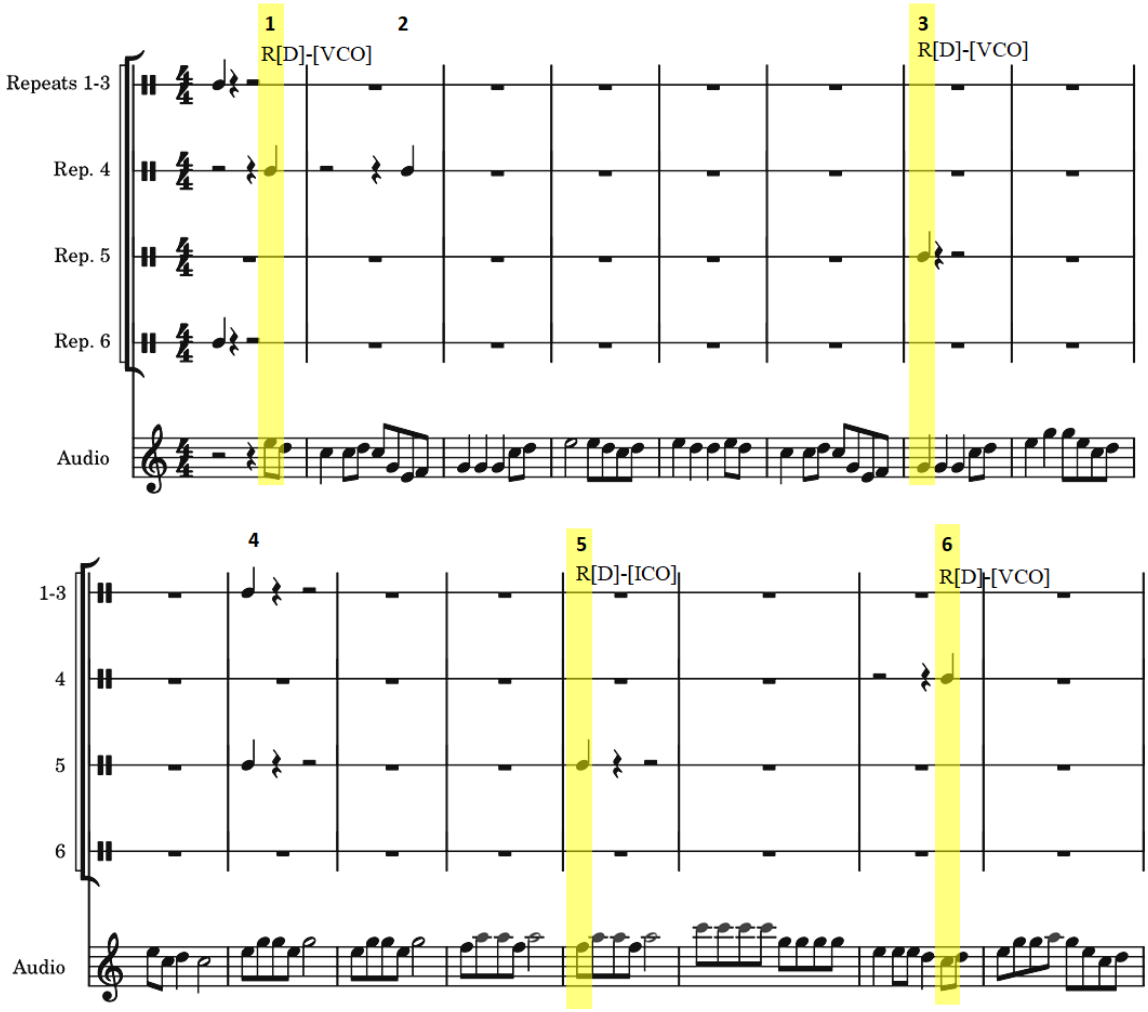


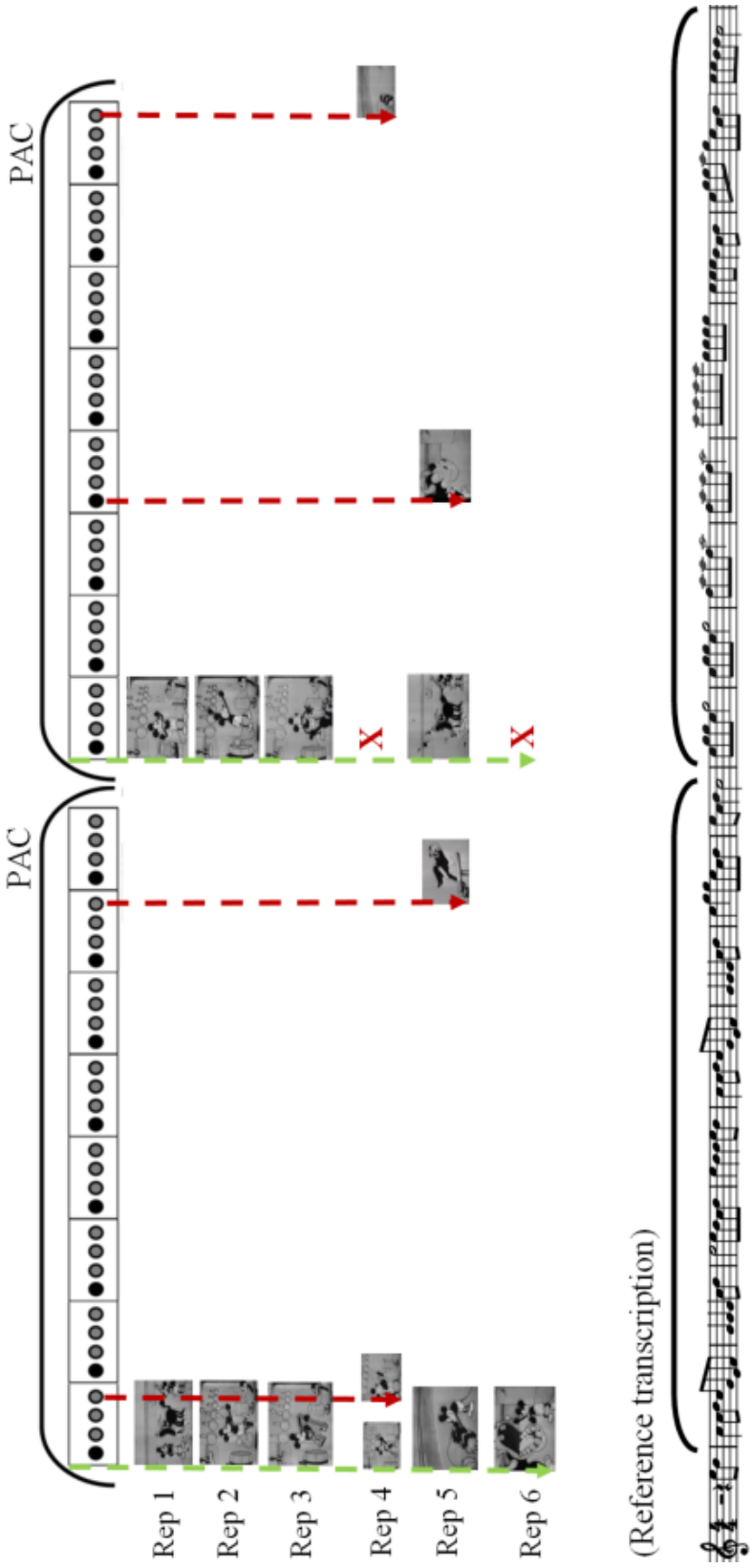
Figure 3.12: [VCO] in “Turkey in the Straw” from *Steamboat Willie* (1928).[199]

3.12a.



[198] “Love is an Open Door”: https://youtu.be/kQDw88hEr2c?si=V-ixJYjIKXOb64Sf. Excerpt begins at 0:06.
[199] *Steamboat Willie*: https://youtu.be/BBgghnQF6E4?si=AJdVQ77eQso8C0h9. Excerpt begins 4:22.

3.12b.Visual summary of *Steamboat Willie* correlated to “Turkey in the Straw”



Chapter 4 Figures

Figure 4.1: Metric sync in "Cinema" from *Hatsune Miku: Colorful Stage!*. Screenshots by the author.[200]

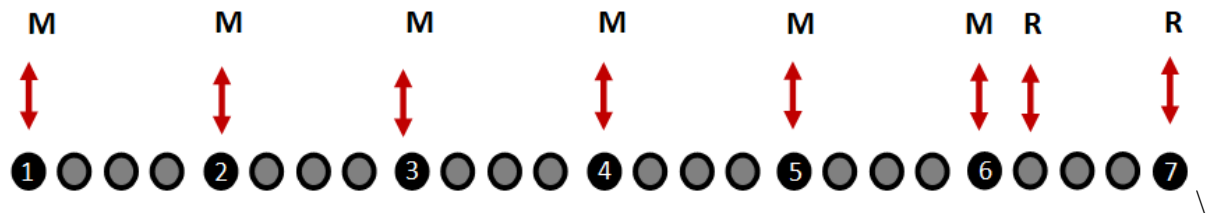


Image progression:

Events:
1. Shot change to An
2. Zoom out to Kohane
3. Zoom out to Akito
4. Zoom out to Toya
5. Zoom out to Kaito
6. Swipe transition to Akito; 6/1: Screen flashes with shot change to Akito

[200] "Cinema": https://youtu.be/8m1cUgLepmY?si=Pgkg6wg69kS9nKWi. Excerpt begins 1:19.

Figure 4.2: Hypermetric sync in a passage from Disney's *Flamingos* (2000).

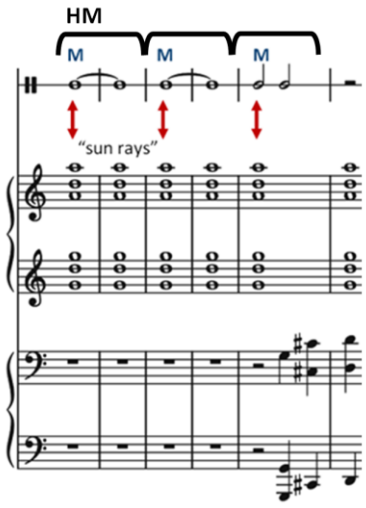


Figure 4.3: R[S]-[VCO] "transforms" (elevates) to M[S]-[VCO] in "Cinema" (2021) from *Hatsune Miku: Colorful Stage!* (2020).

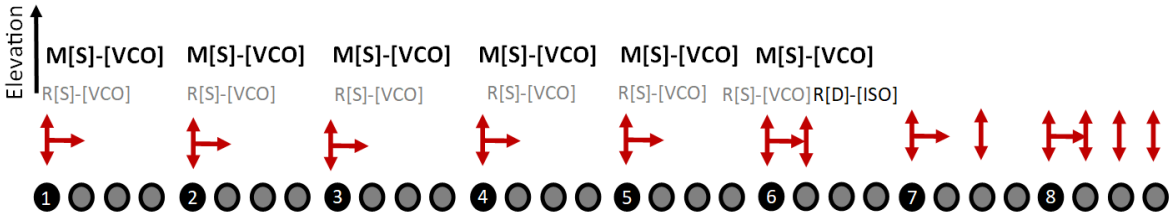


Figure 4.4: (Hyper)metric sync in "Turkey in the Straw," *Steamboat Willie*, repetitions 1–3.

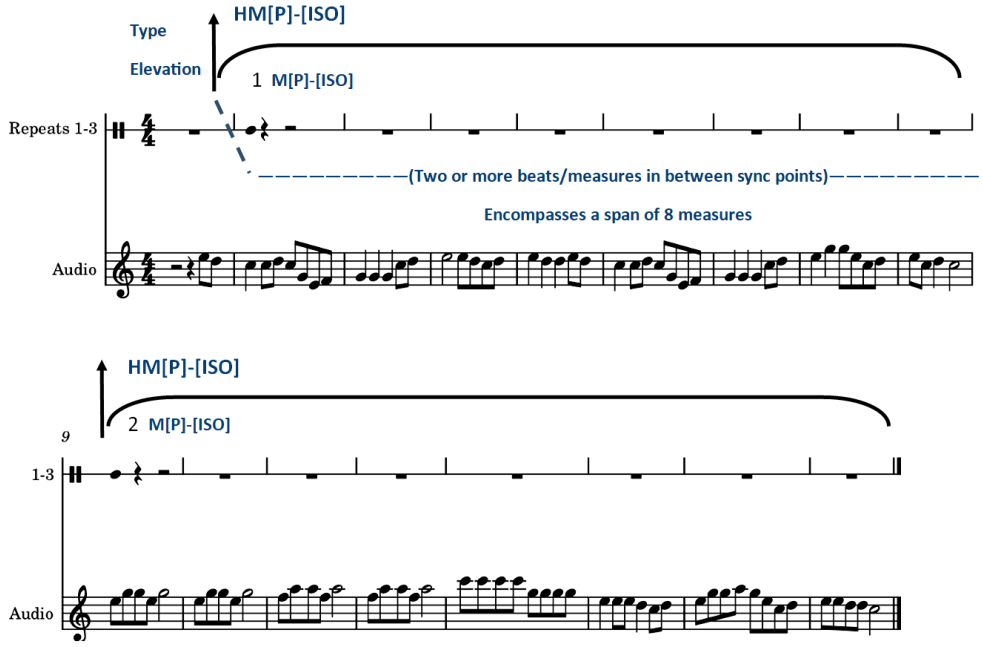


Figure 4.5: Sync rates in “Turkey in the Straw,” repetitions 1–6. Transcription by the author.

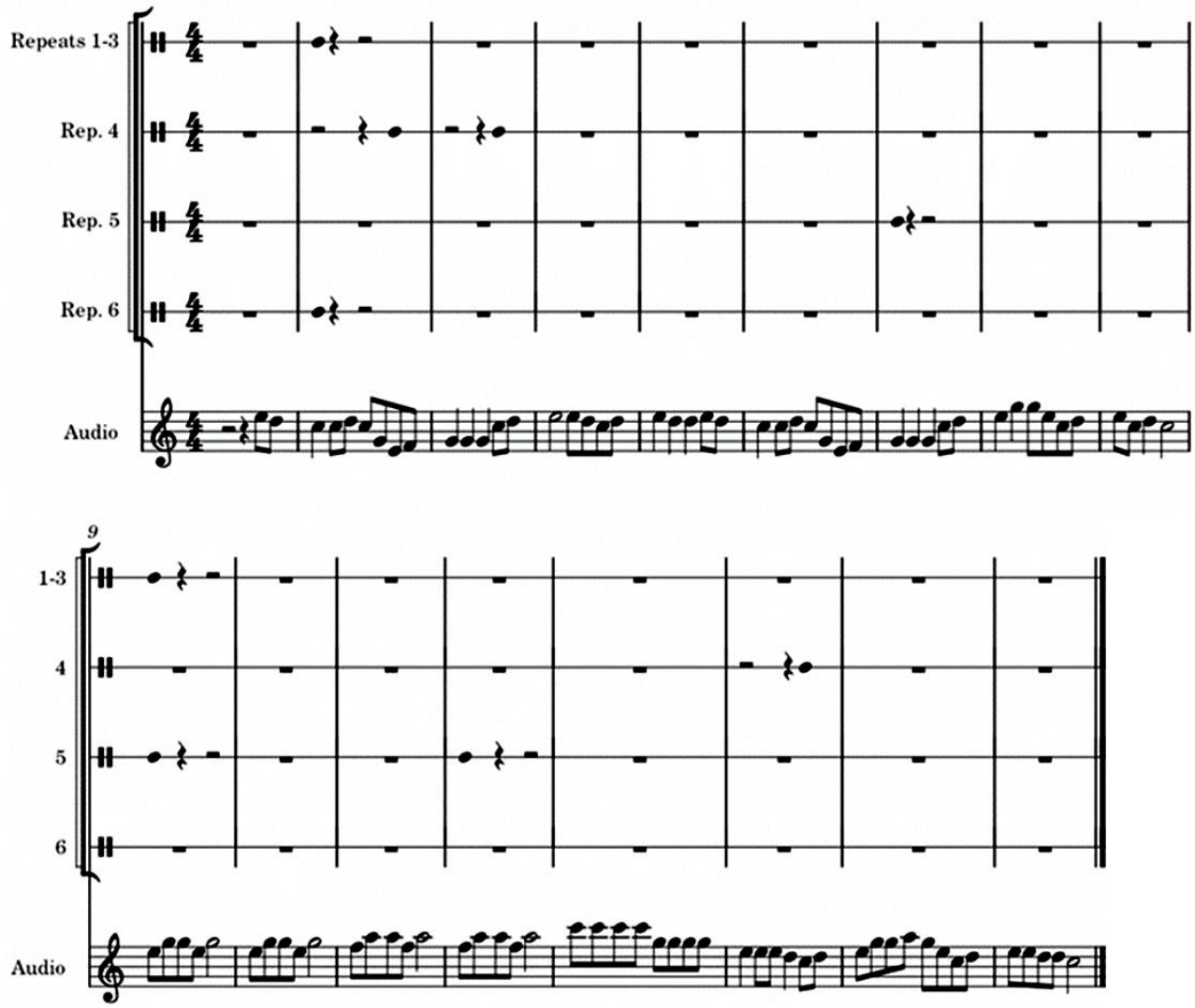
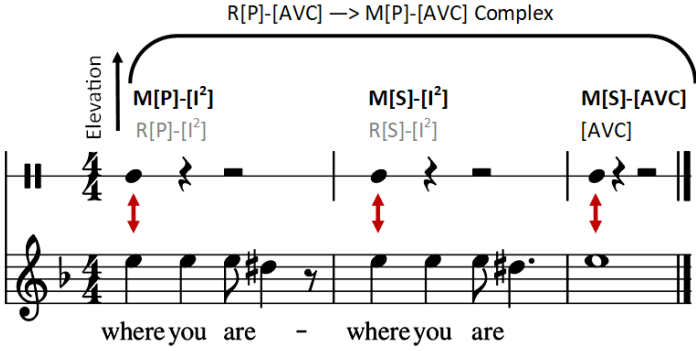
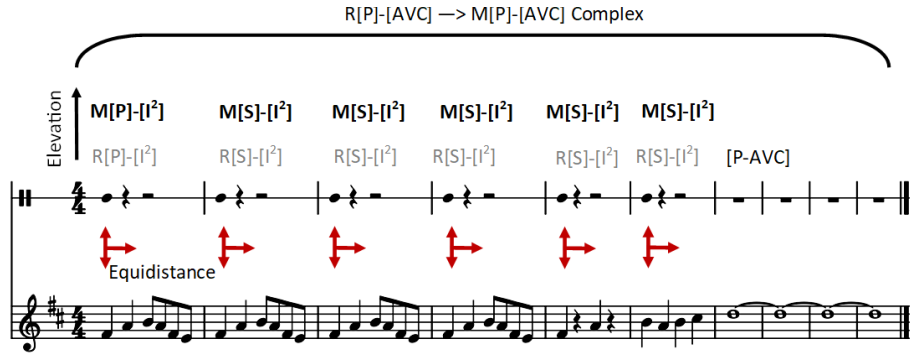


Figure 4.6: Examples of metric and non-metric sync elevation at [AVC] complexes.

4.6a. “Where You Are,” Moana (2016)



4.6b. “Honor to Us All,” *Mulan* (1998)



4.6c. “Poor Unfortunate Souls,” *The Little Mermaid* (1989)

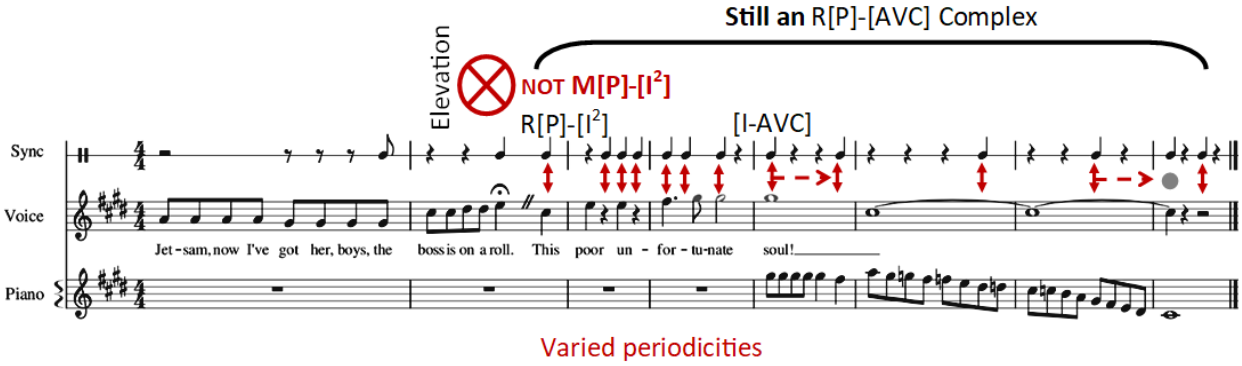


Figure 4.7: Hypermetric sync in a passage from *Flamingos*, from *Fantasia 2000*.

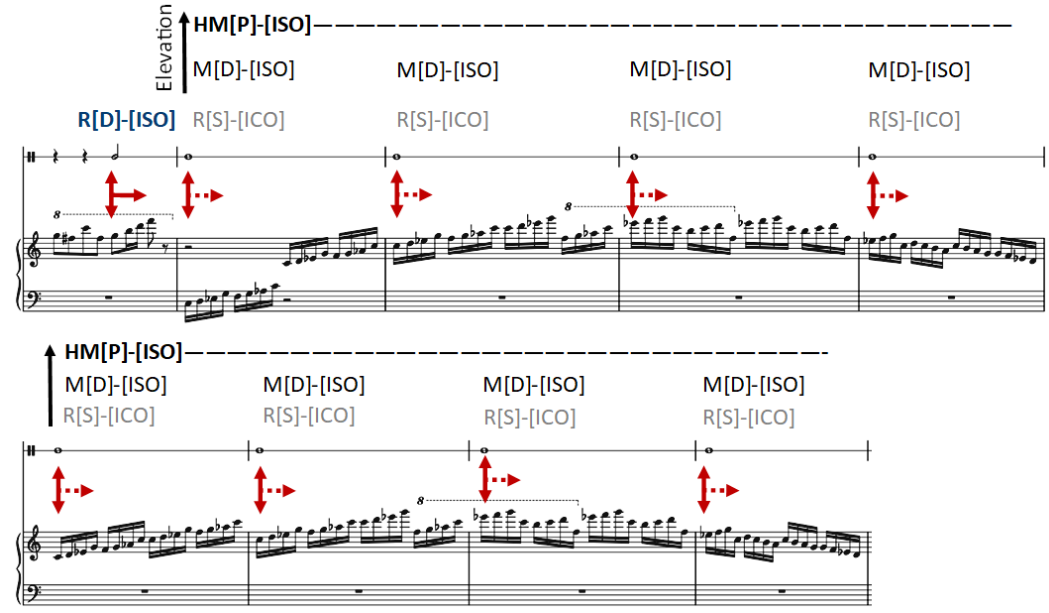


Figure 4.8: Pyramidic relationship among R-, M-, and HM-type sync points.

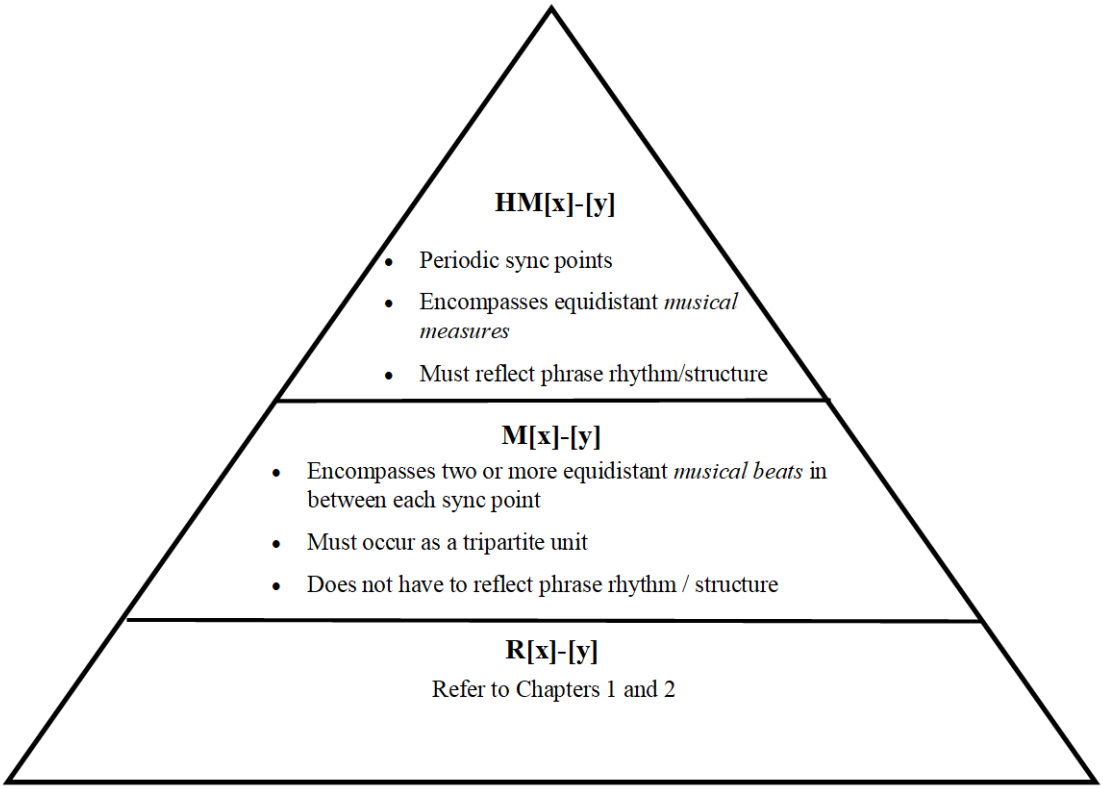


Figure 4.9: Tripartite tone projection unit, adapted from Hasty (1997).

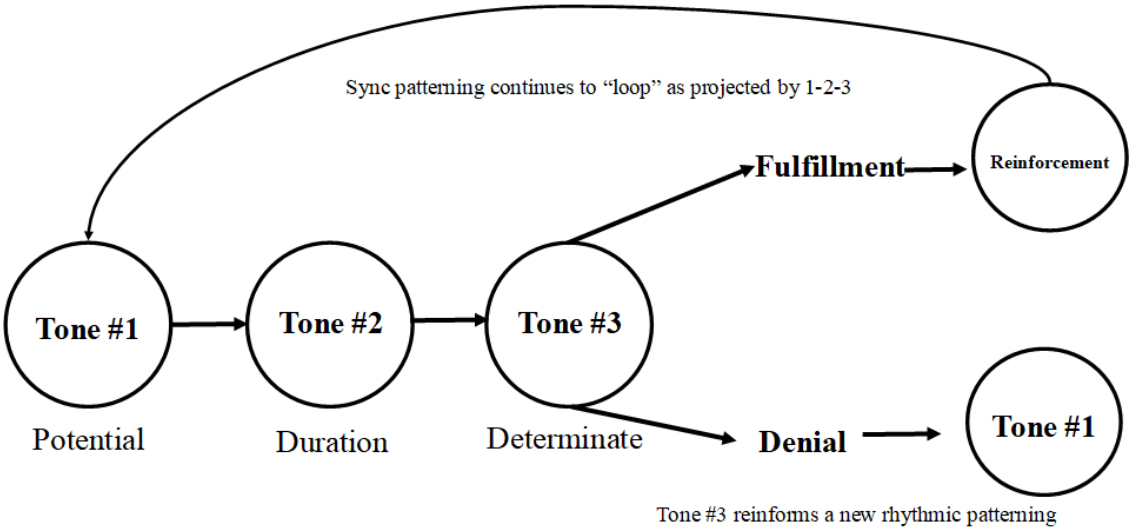
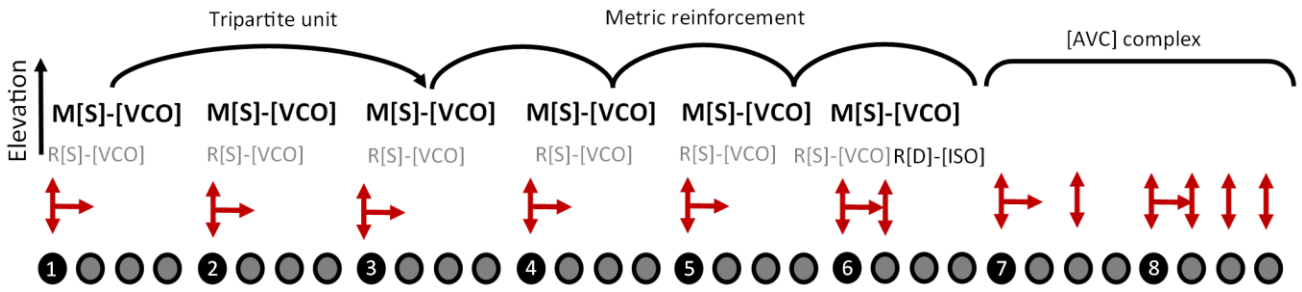


Figure 4.10: Projection applied to [AVC]. "Cinema," *Hatsune Miku: Colorful Stage!* (2020).

4.10a.



4.10b.

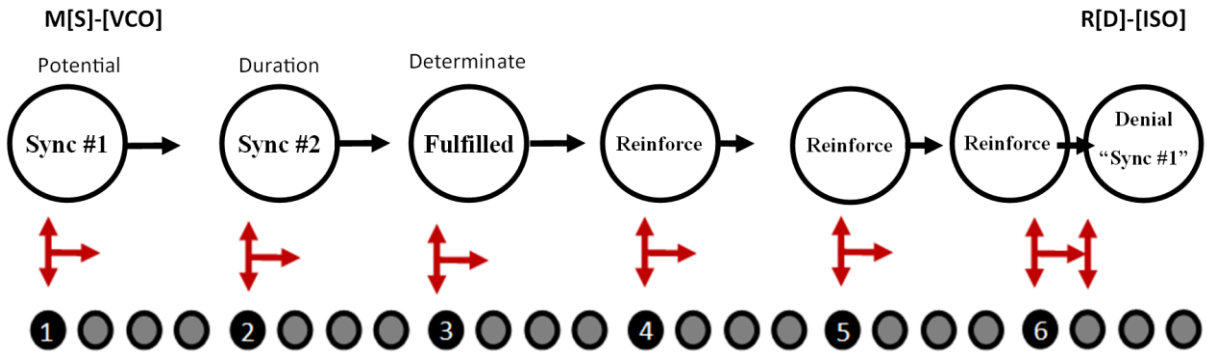


Figure 4.11: Projection unit applied to the opening of *Flamingos* from Disney's *Fantasia 2000*.

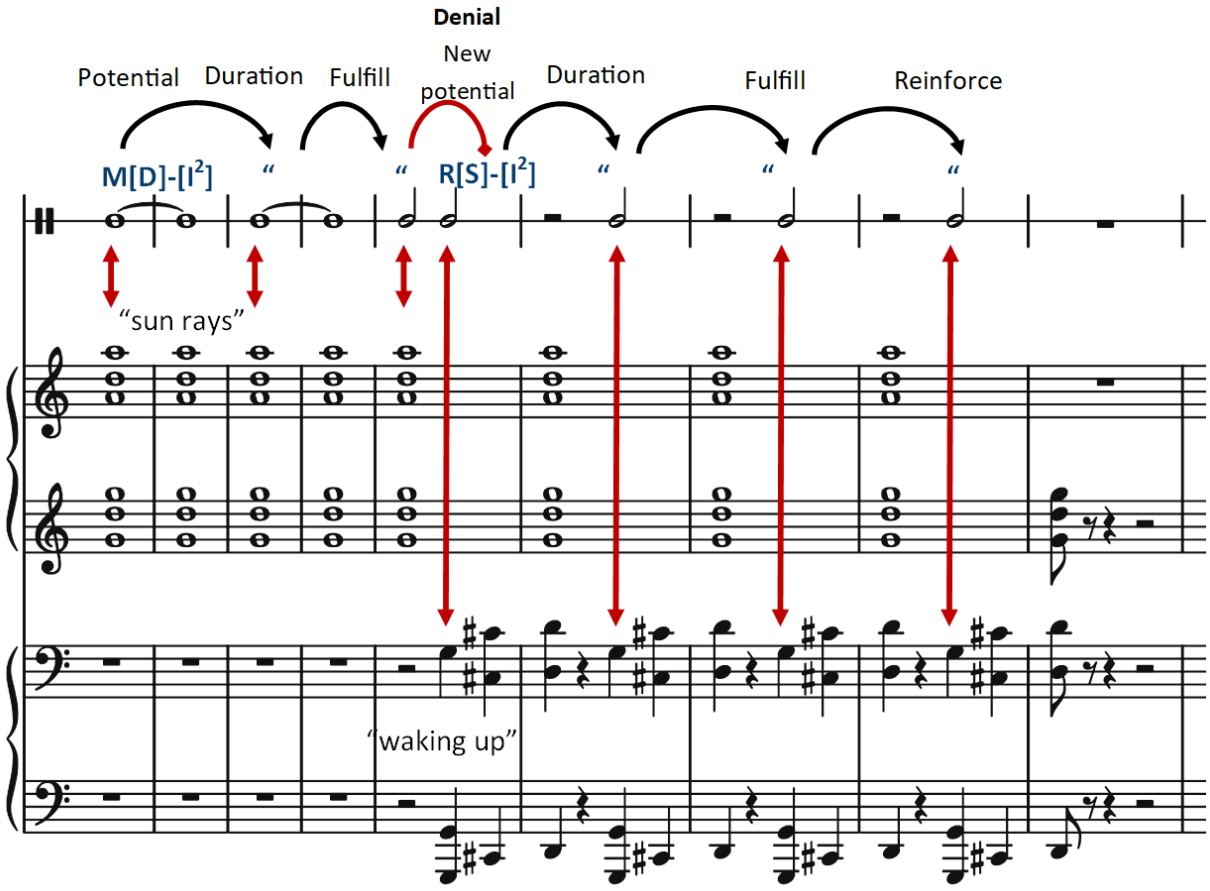
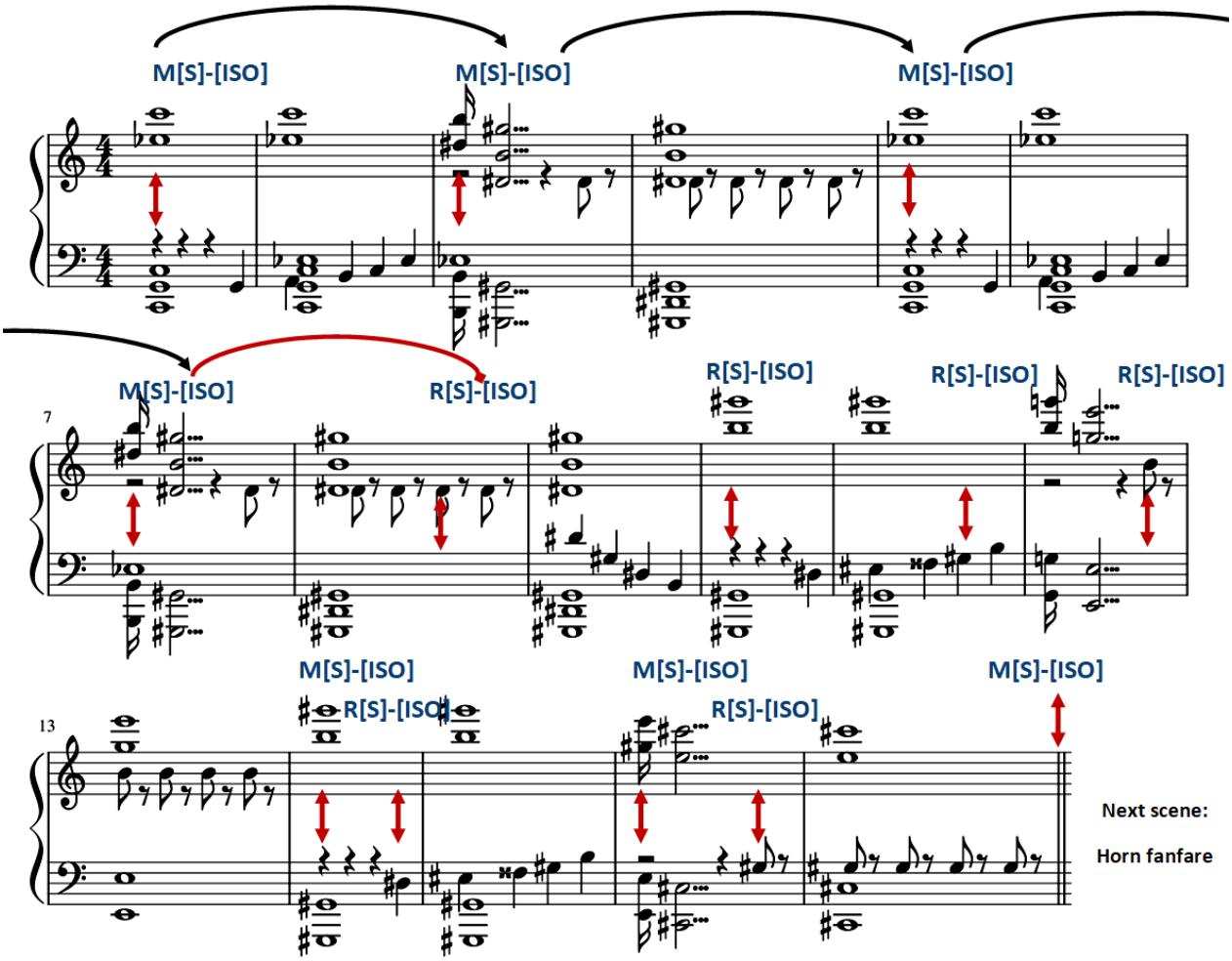
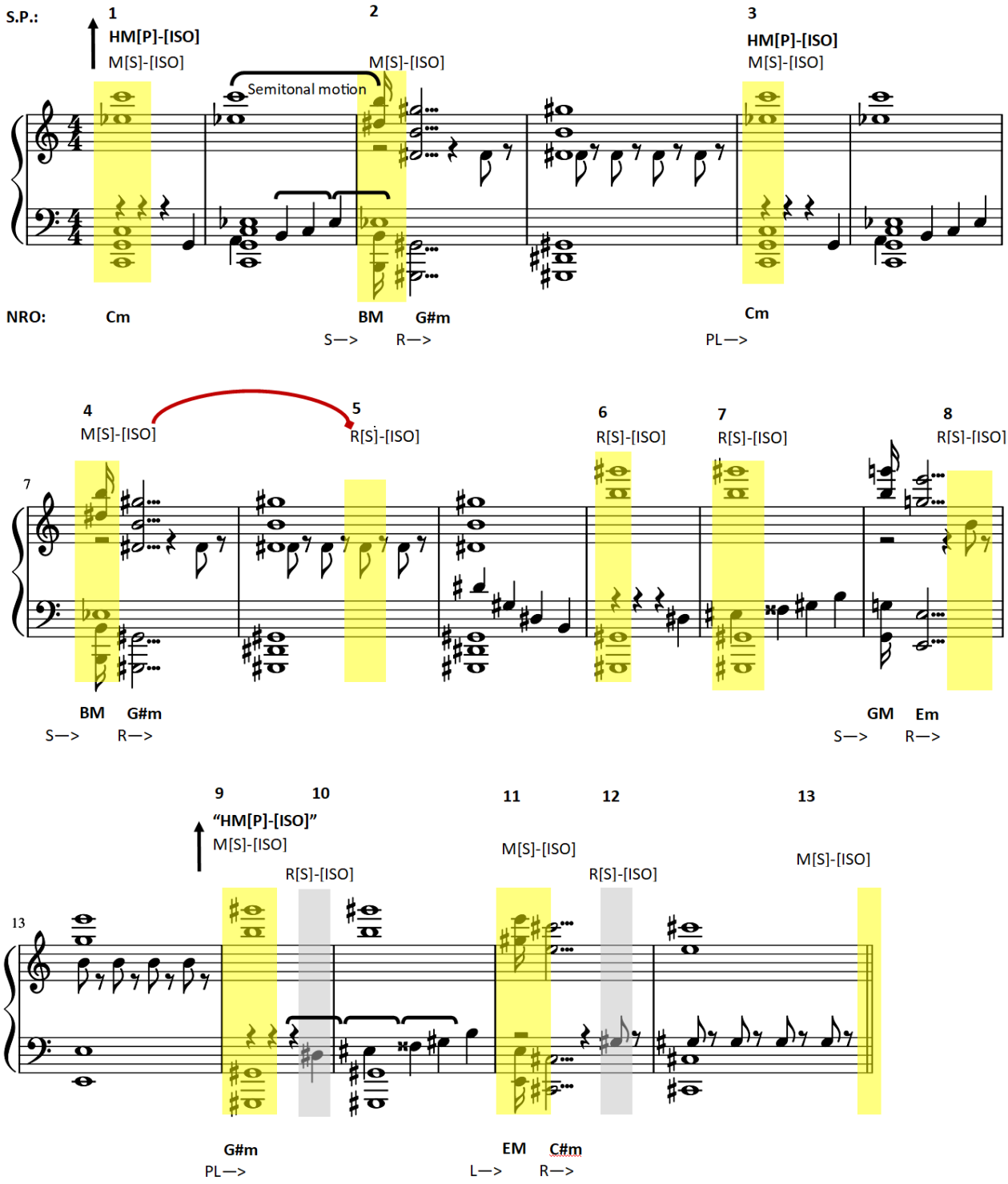


Figure 4.12: Metric sync and projection applied to the "Battle on the Ice" cue from *Alexander Nevsky* (1938).[201]



[201] The transcription refers to this version of the film: Fix Me a Scene, "Alexander Nevsky (1938) — Battle On The Ice," YouTube video, https://youtu.be/IcPixaWL2Pg?si=jhsI-eePg7_hj8YS.

Figure 4.13: “Battle on the Ice” cue with added harmonic analysis, *Alexander Nevsky* (1938).



Chapter 5 Figures

Figure 5.1: Visualization of sync point projection.

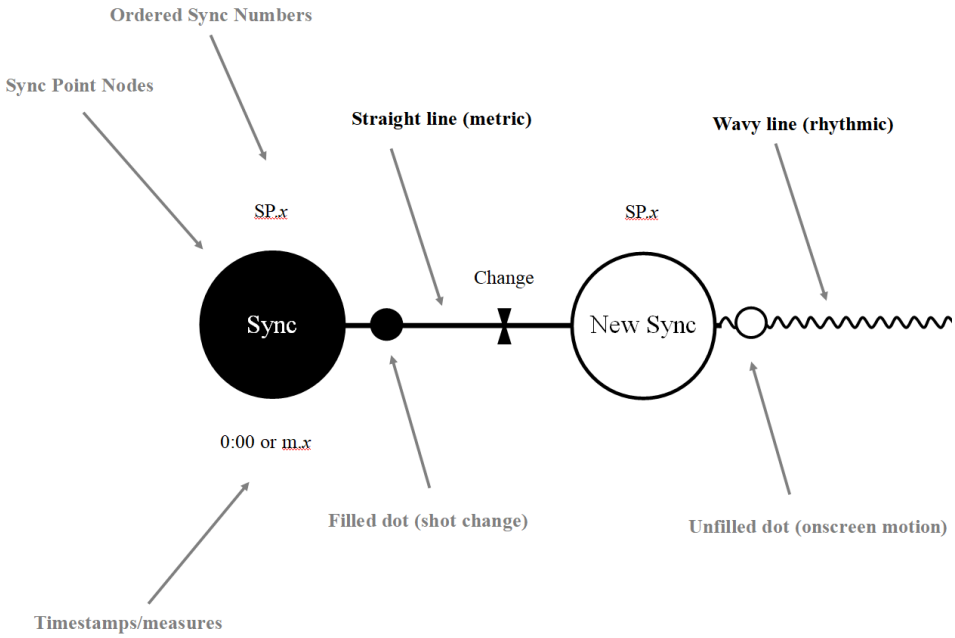
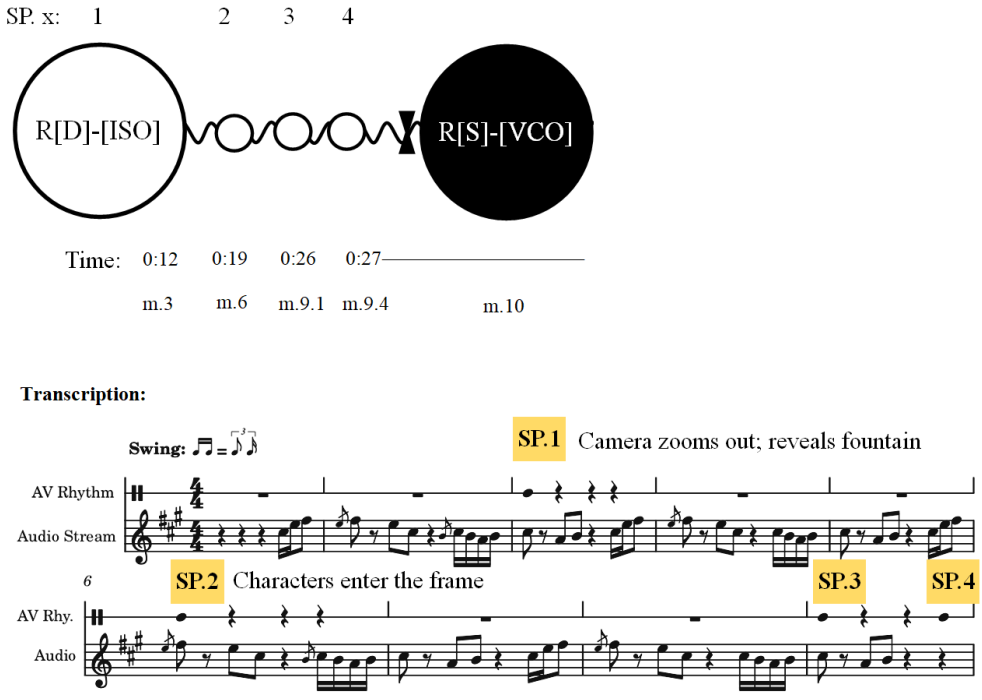


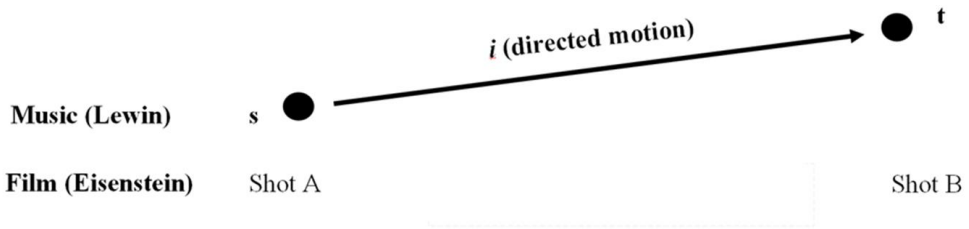
Figure 5.2: A sync train with information. Transcription of "Close to Gray" by the author.[202]



[202] Watch "Close to Gray" here: https://youtu.be/8m1cUgLepmY?si=GmO1evPFbLgF4A18.

Figure 5.3: Correlation between Lewin's transformational attitude and Eisenstein's montage, as well as Pudovkin's filmic space.

5.3a.



5.3b.

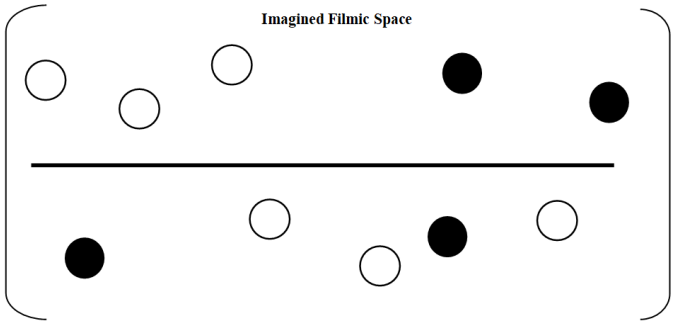


Figure 5.4: Hierarchies within a sync train.

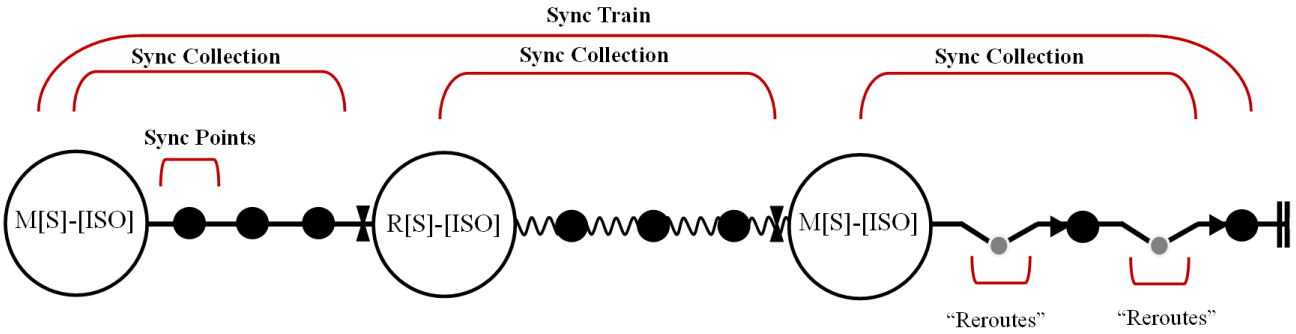


Figure 5.5: Sync train for “Battle on the Ice” cue from *Alexander Nevsky* (1938). Transcription by the author.

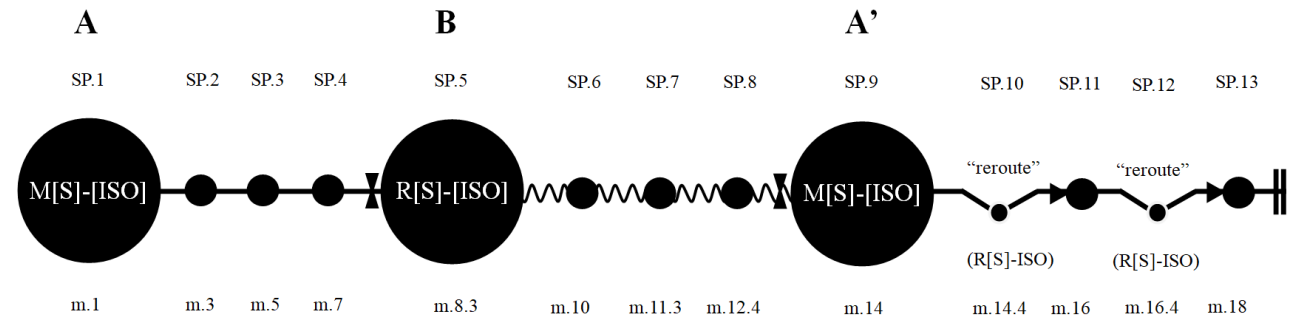


Figure 5.6: Temporal data and trends found in sync trains.

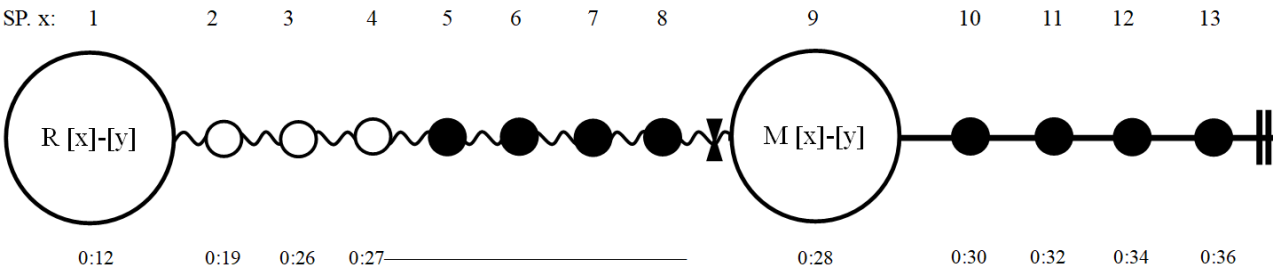
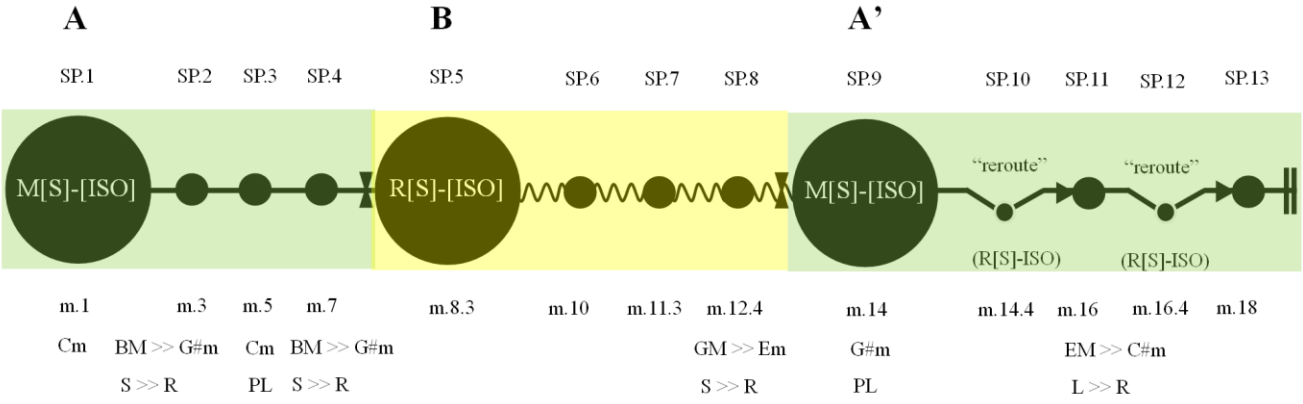


Figure 5.7: Sync train for “Battle on the Ice” cue, with color shading. Transcription by the author.

5.7a.



5.7b.

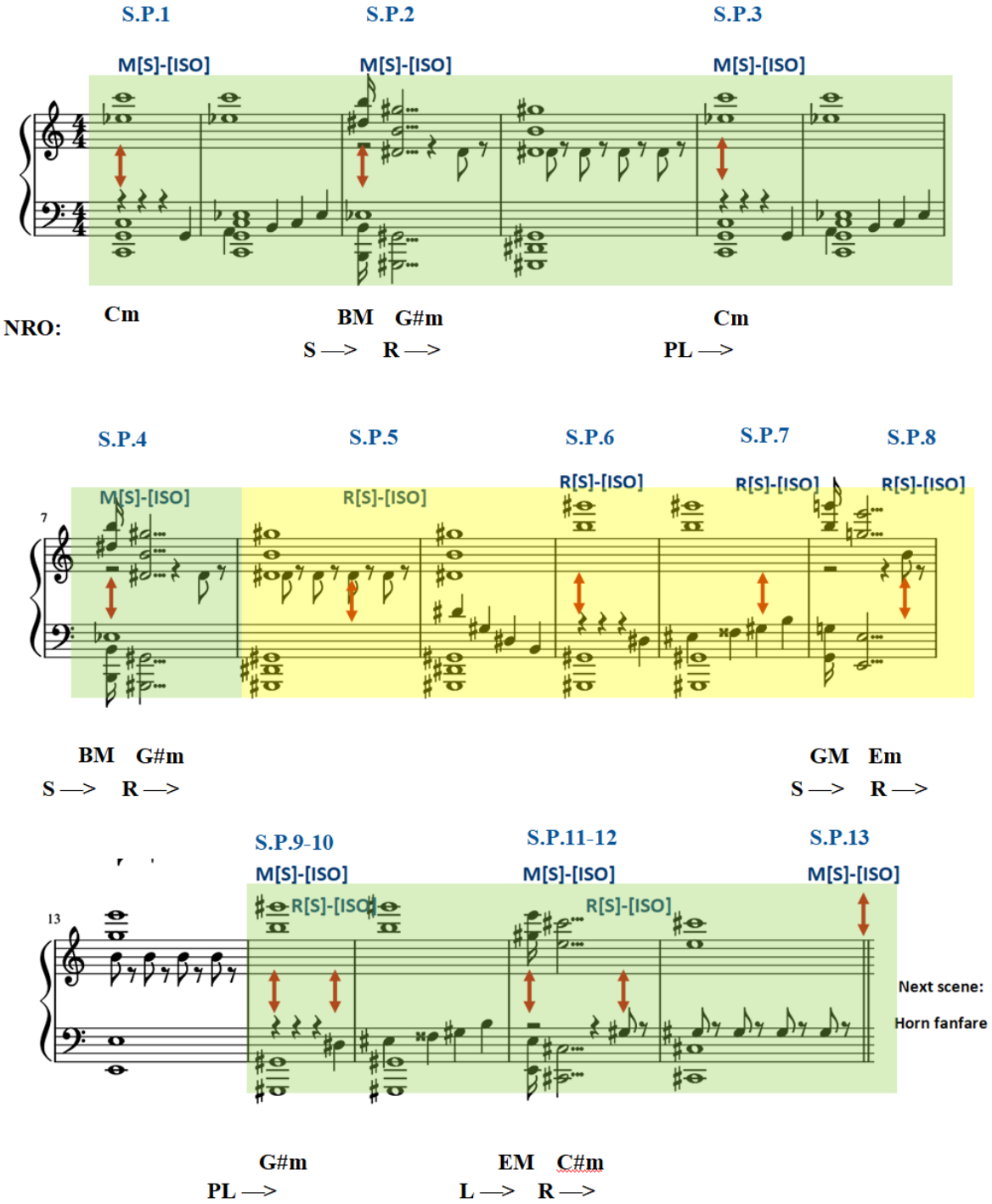
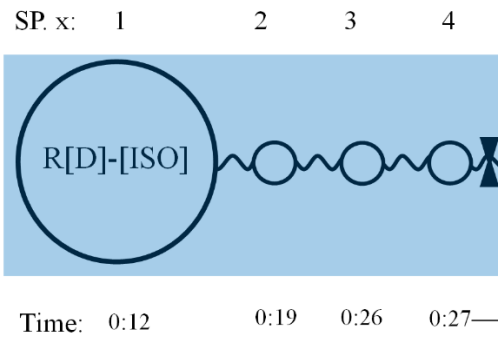

S.P.1
S.P.2
S.P.3
M[S]-[ISO]
M[S]-[ISO]
M[S]-[ISO]
NRO:
Cm
BM G#m
S —> R —>
Cm
PL —>
S.P.4
S.P.5
S.P.6
S.P.7
S.P.8
M[S]-[ISO]
R[S]-[ISO]
R[S]-[ISO]
R[S]-[ISO]
R[S]-[ISO]
7
BM G#m
S —> R —>
GM Em
S —> R —>
S.P.9-10
S.P.11-12
S.P.13
M[S]-[ISO]
M[S]-[ISO]
M[S]-[ISO]
R[S]-[ISO]
R[S]-[ISO]
13
Next scene:
Horn fanfare
G#m
PL —>
EM C#m
L —> R —>

Figure 5.8: Example of *Aurality* in a sync train graph.

5.8a



5.8b

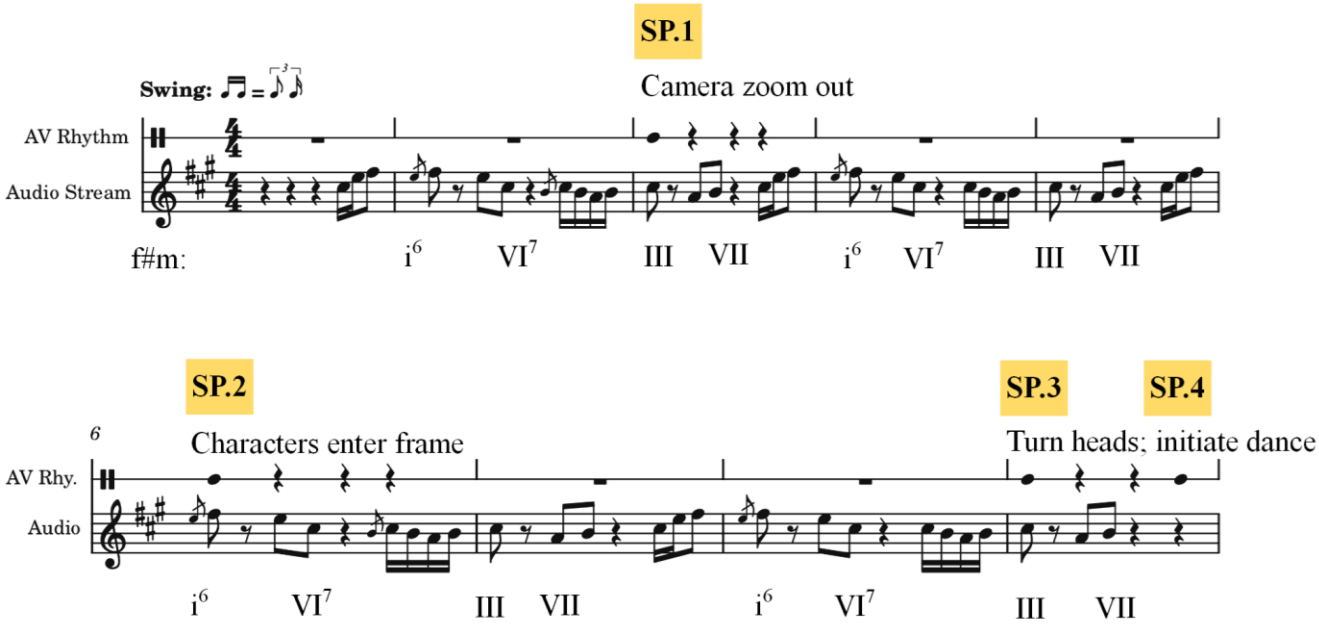


Figure 5.9: Example of *Melodic Totality* in “Under the Sea” from *The Little Mermaid* (1989).

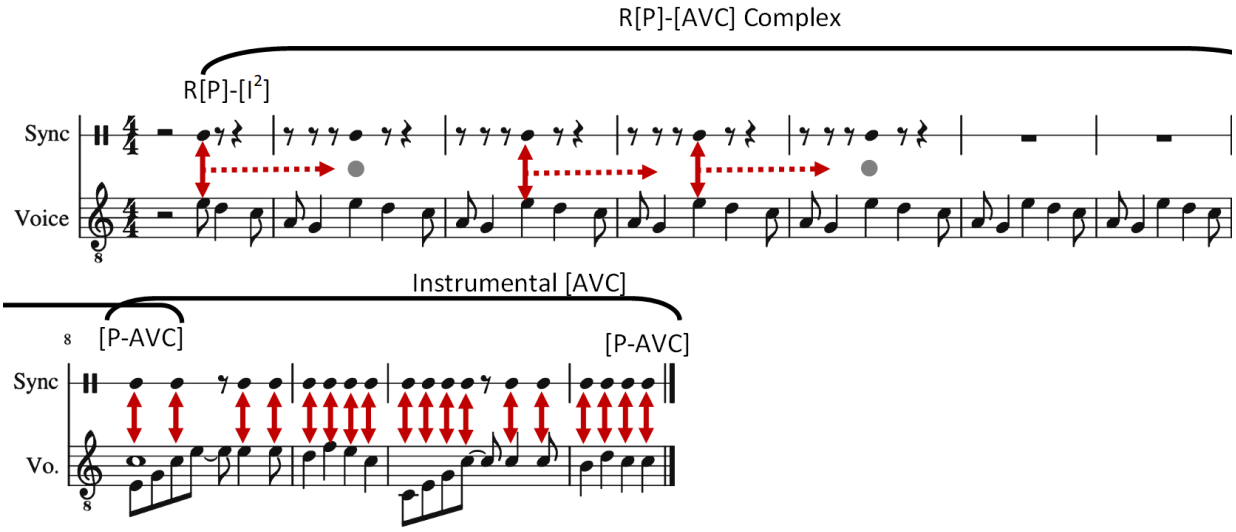


Figure 5.10: Complete Sync Train for “Close to Gray” by Nightcord at 25:00 (2021).

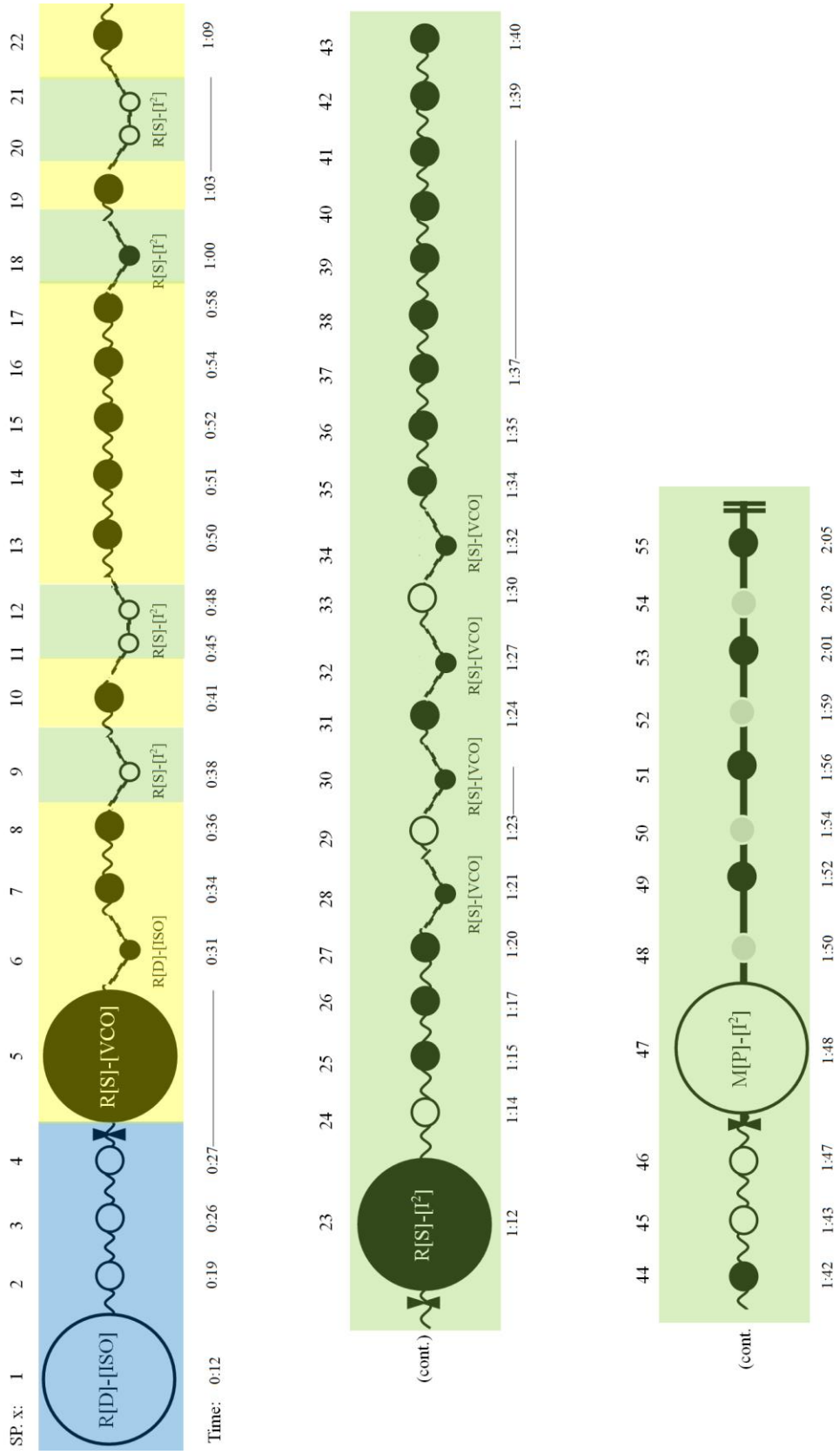
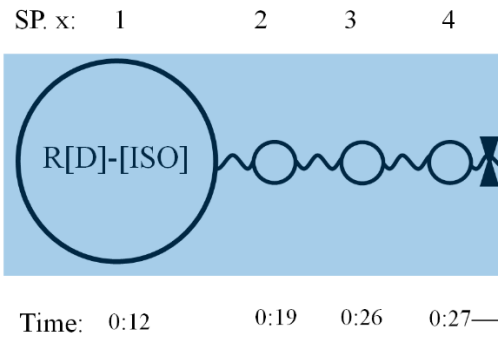
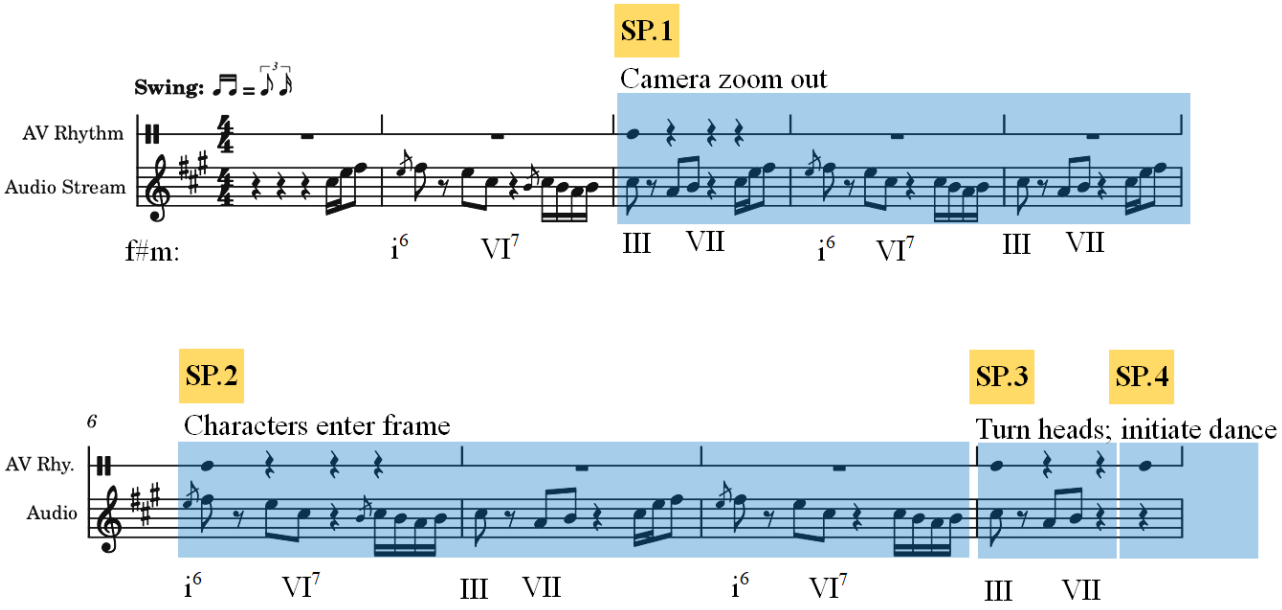


Figure 5.11: Example of *aurality* in the opening of "Close to Gray."[203]

Sync collection:



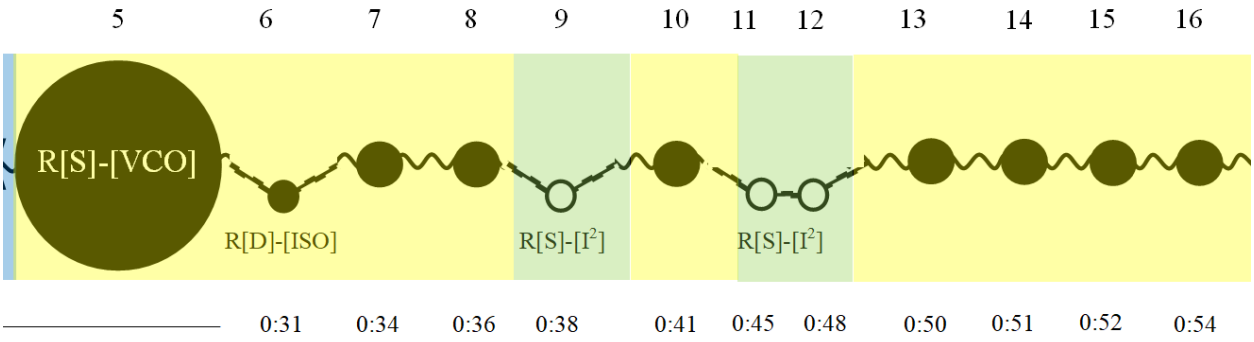
Transcription:



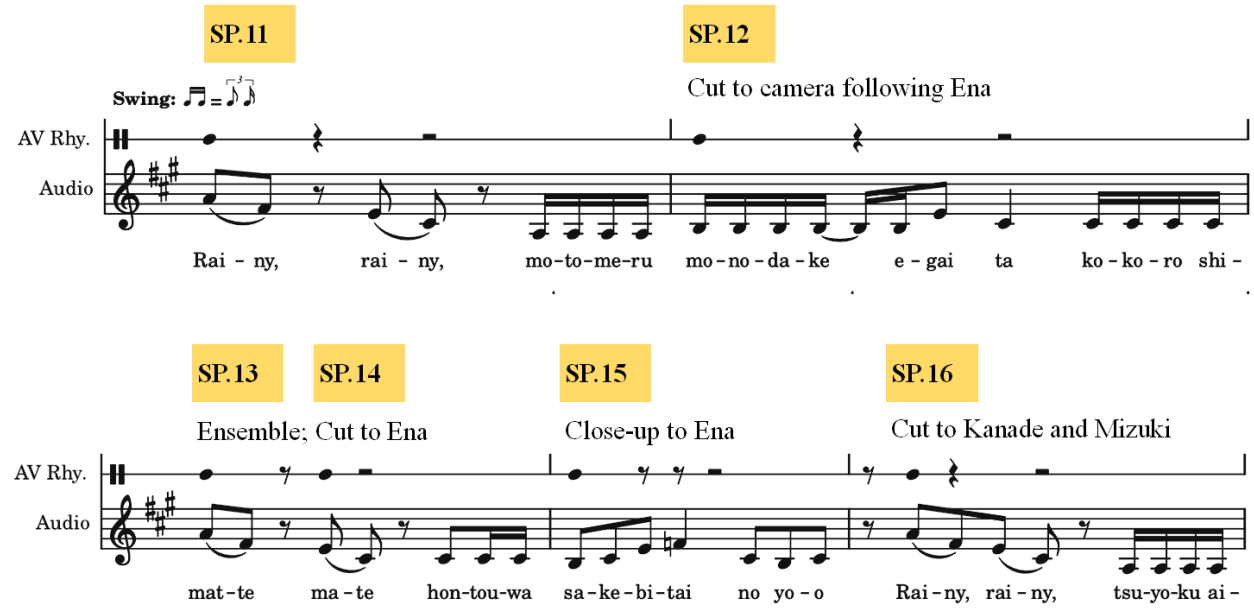
[203] See the exact moment here: https://youtu.be/3054q5rA5JQ?si=lIGItOvjktplmKw6

Figure 5.12. Example of *visuality* in "Close to Gray." Transcription by author.[204]

Sync collection:



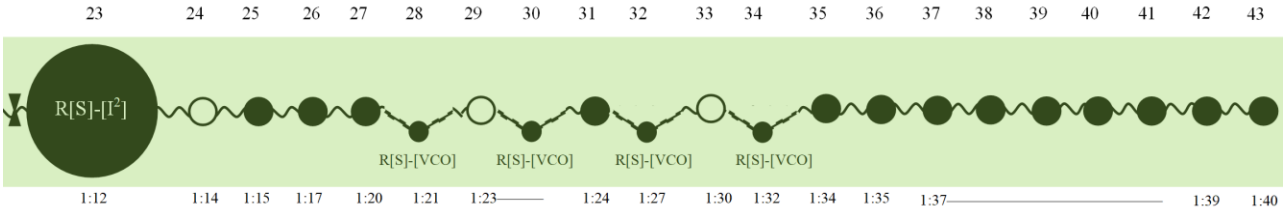
Transcription:



[204] See the exact moment here: https://youtu.be/3054q5rA5JQ?si=ehBWNGcuhzoht9L0&t=45

Figure 5.13: Example of *totality* in "Close to Gray."[205]

Sync collection:



Transcription:

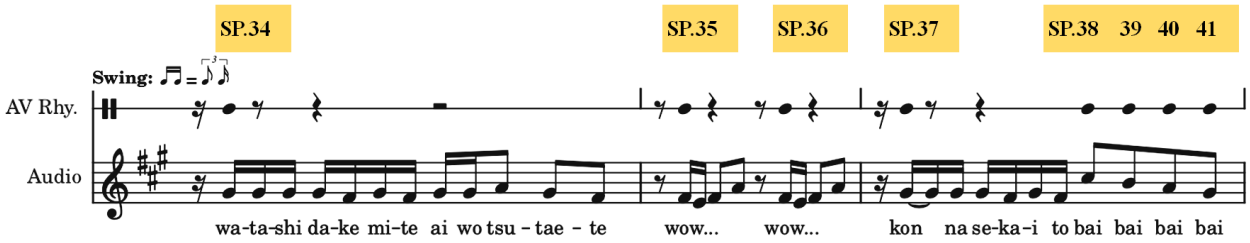
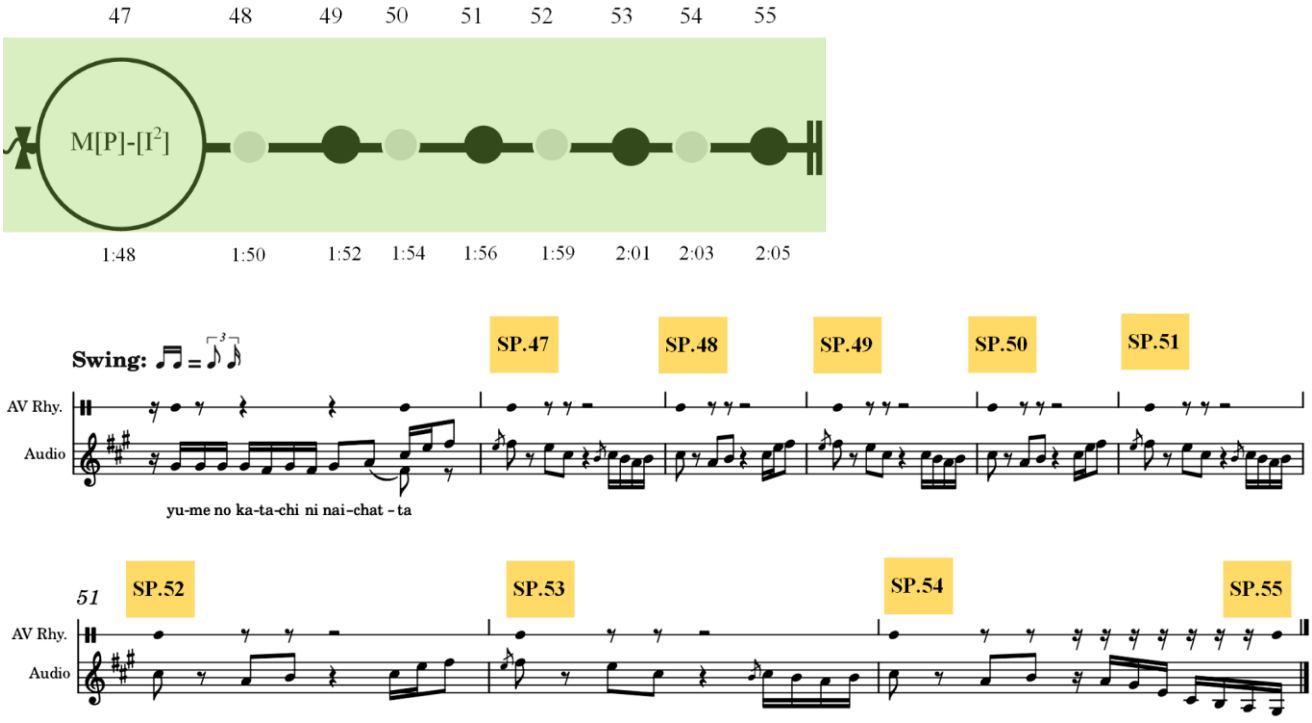
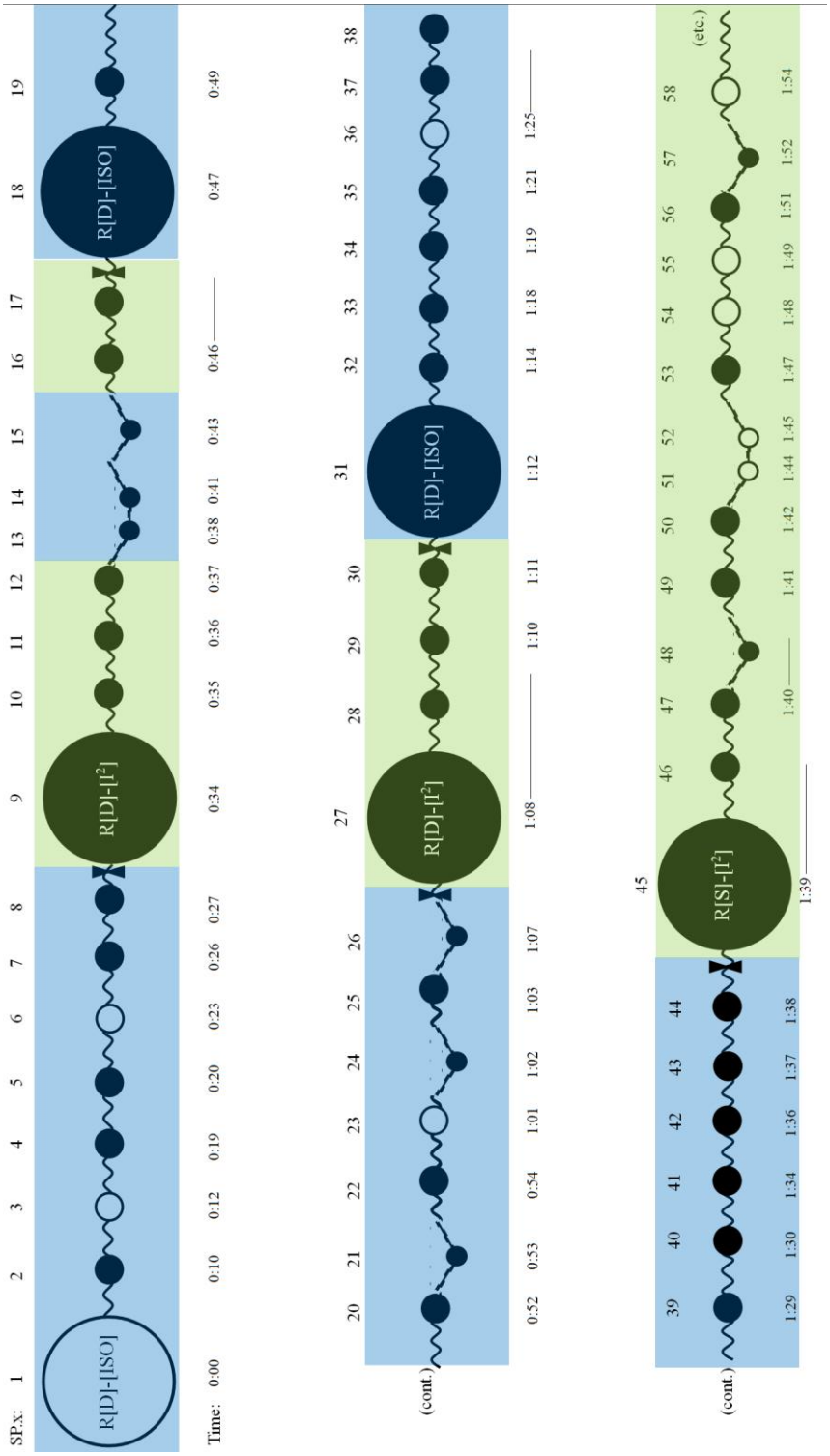


Figure 5.14: Metric sync in the closing passages of "Close to Gray." Transcription by author.[206]



[205] See the exact moment here: https://youtu.be/3054q5rA5JQ?si=6TLzKD79ngMBK4WN&t=92
[206] See the exact moment here: https://youtu.be/3054q5rA5JQ?si=W9Bvi7W0m61InfiR&t=107

Figure 5.15 Complete Sync Train for “Close to Gray” by Surii (2021).[207]

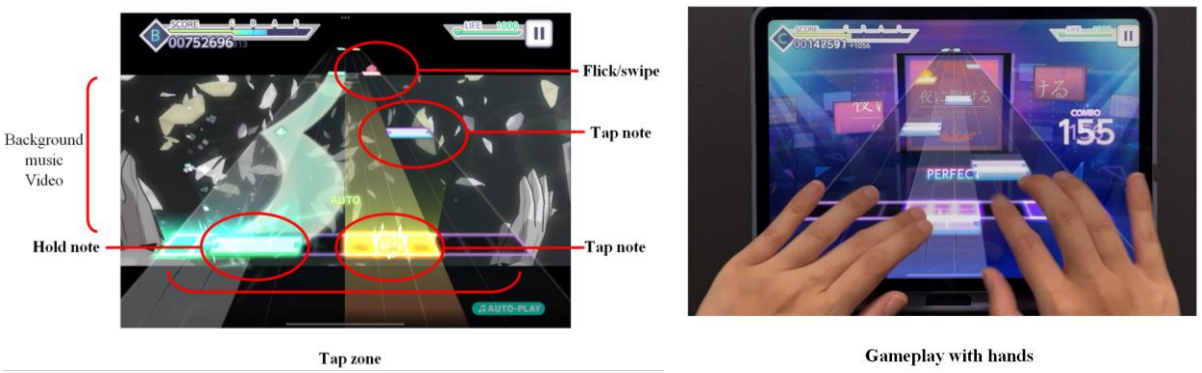


[207] Watch the original music video here: https://www.youtube.com/watch?v=cwLlHFMlBNQ. Last accessed June 2024.

Chapter 6 Figures

Figure 6.1: Rhythm game interface in *Hatsune Miku: Colorful Stage!*

6.1a



6.1b

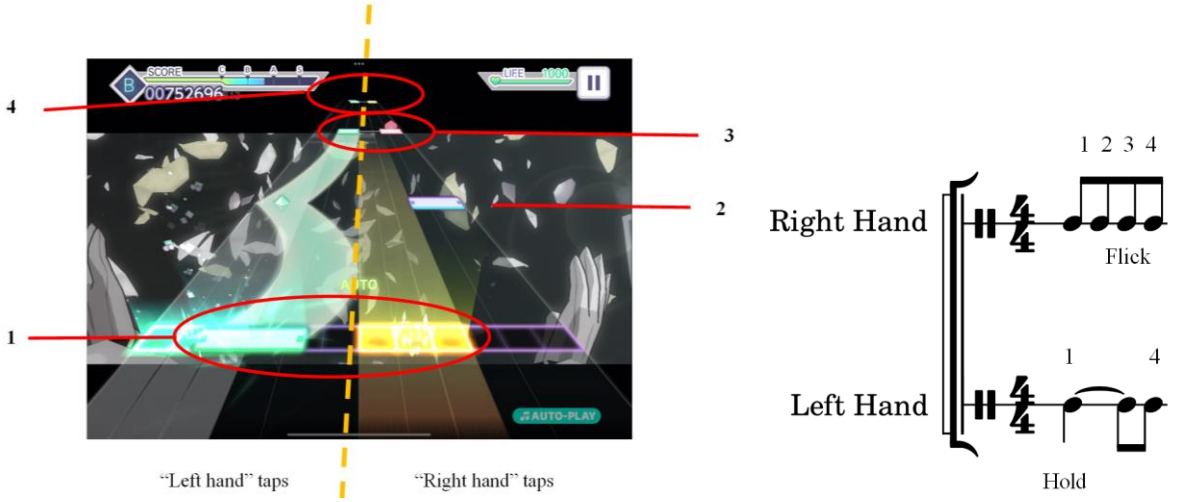


Figure 6.2: Beatmap and synchronicities of "hold" notes. Transcription by the author.

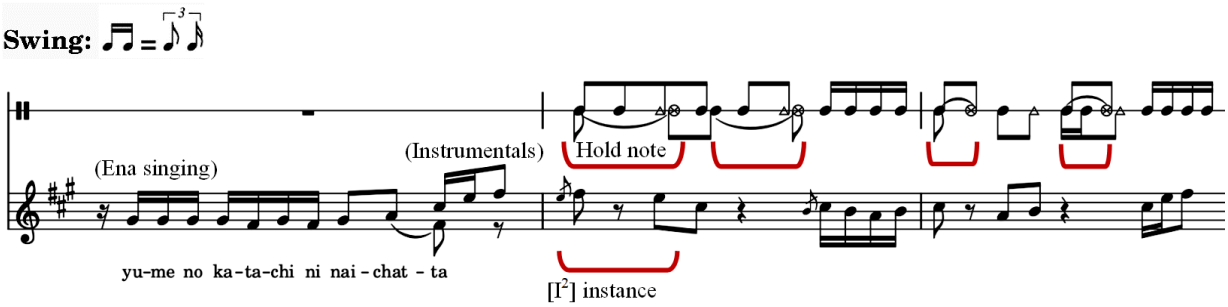


Figure 6.3: Different beatmaps for each difficulty level. Transcription by the author.

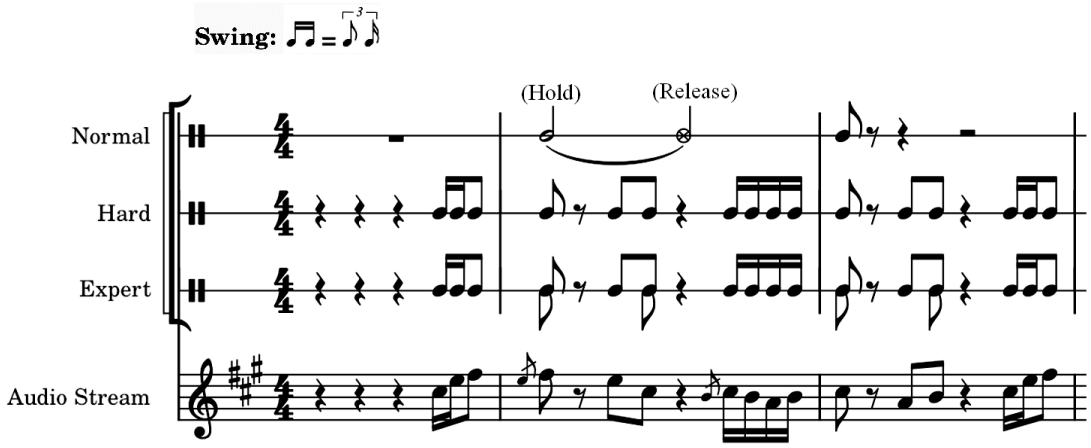


Figure 6.4: Emergent metric relations that are consistent at each difficulty level.

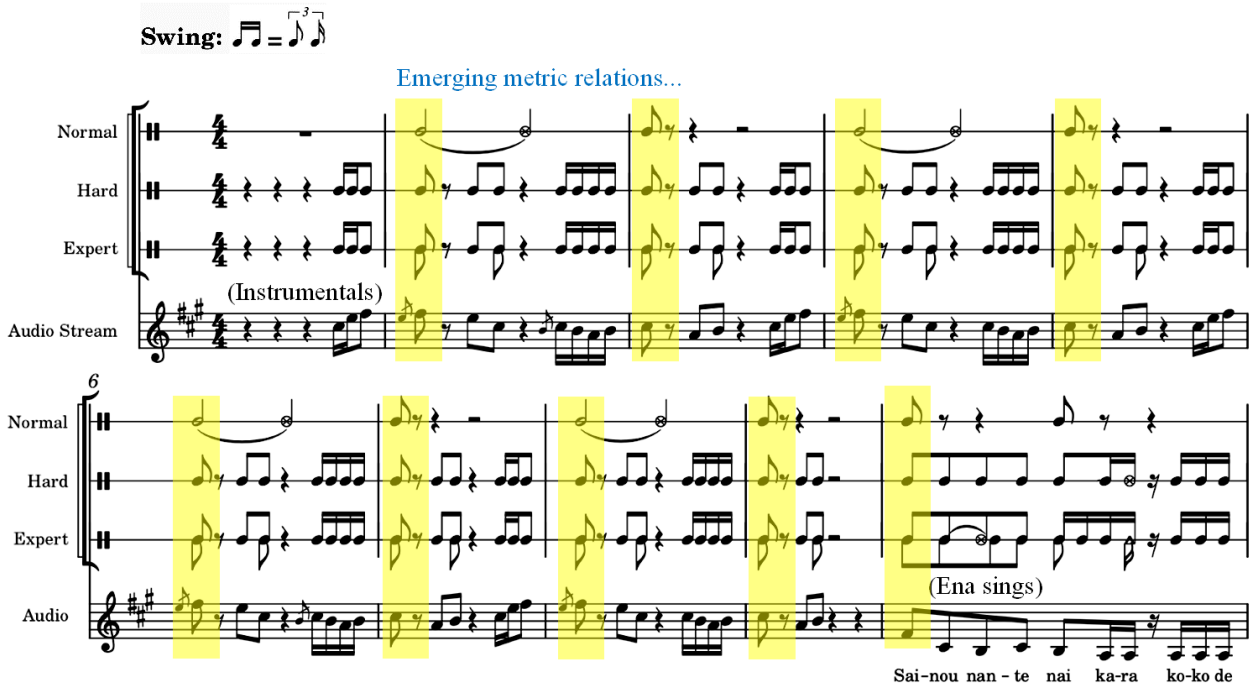


Figure 6.5: How beatmaps organize and frame metric synchronicities.[208]

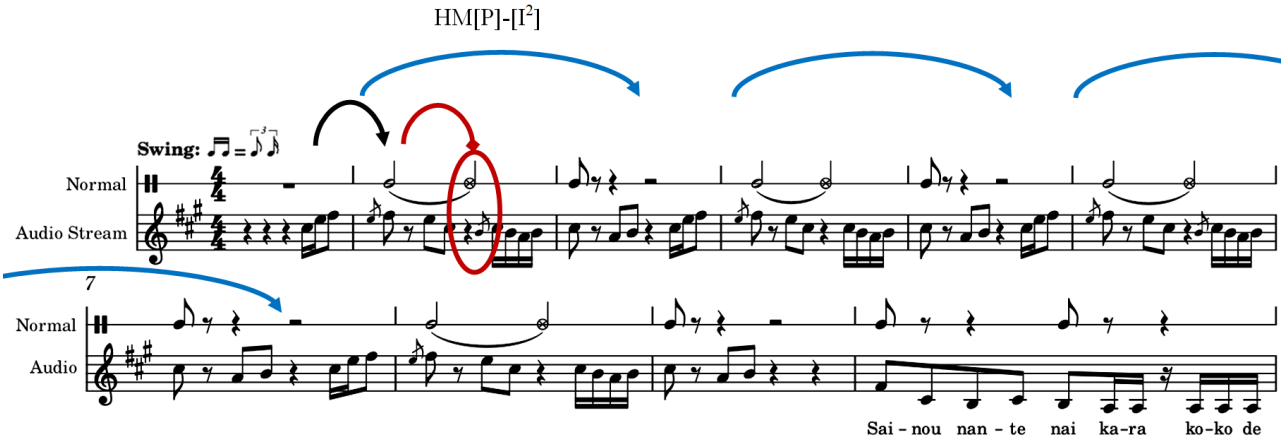
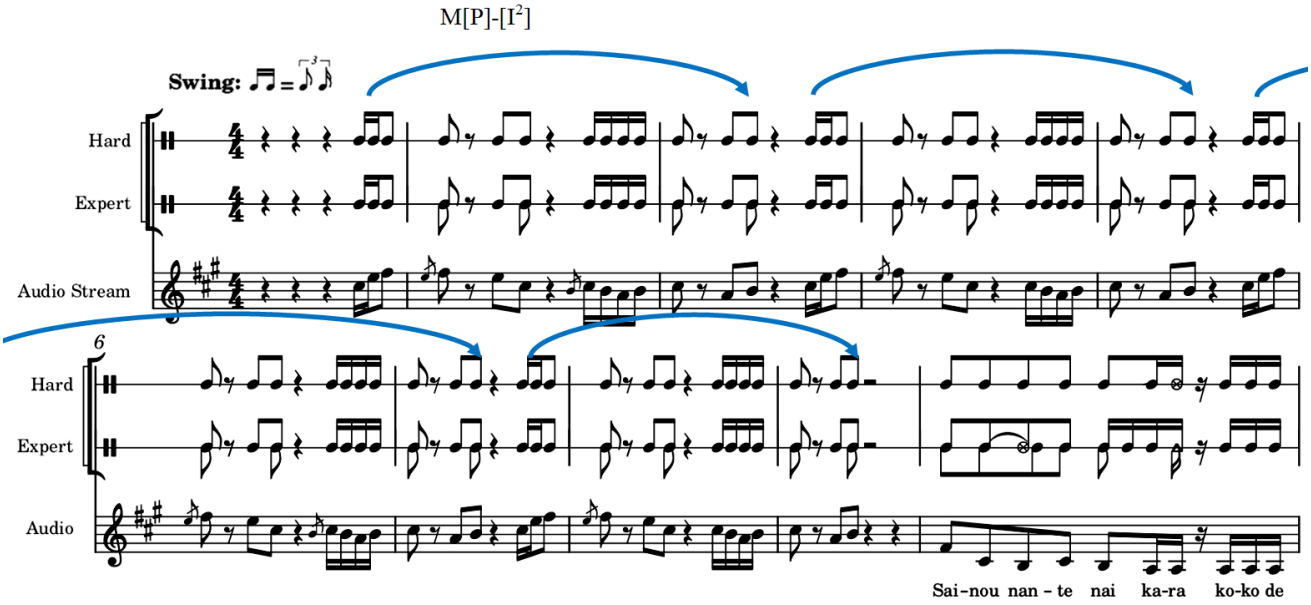


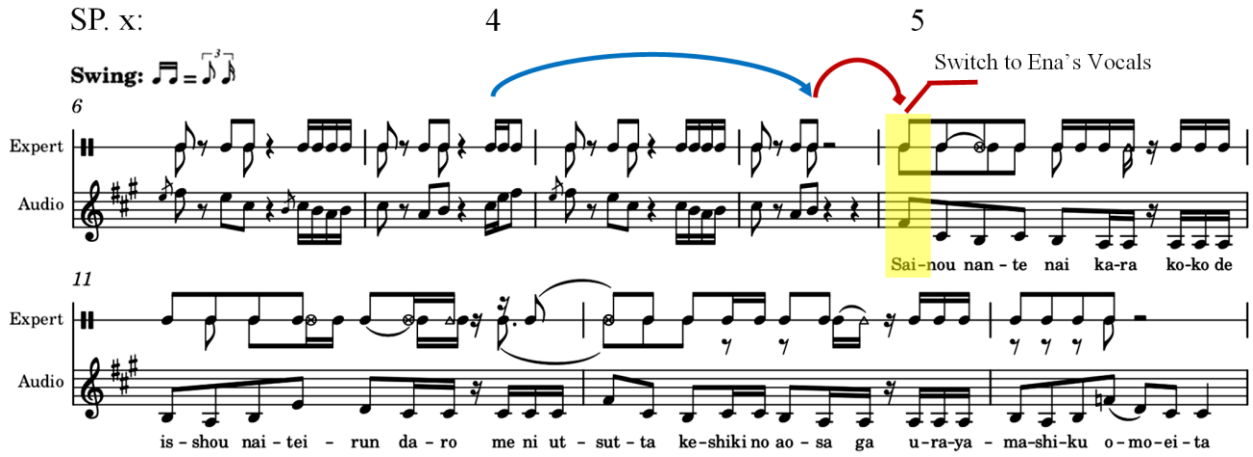
Figure 6.6 New grouping units formed at higher difficulties.



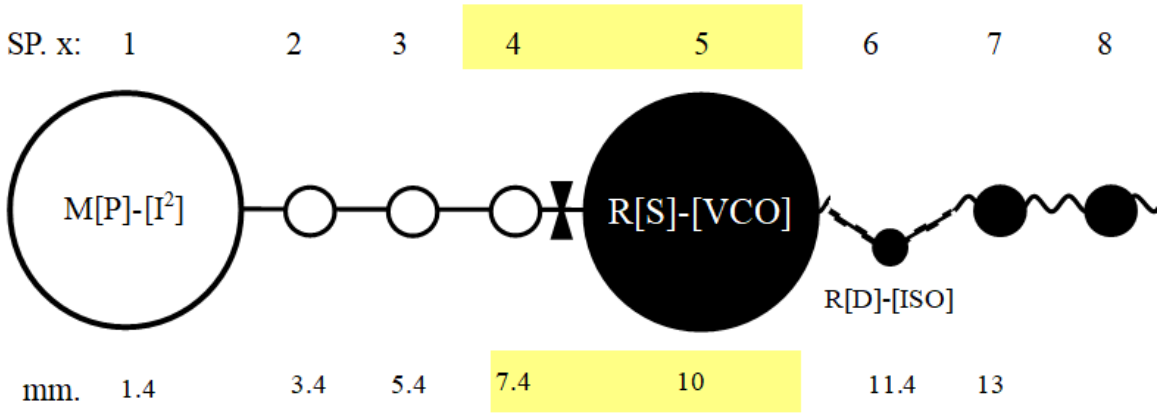
[208] Example gameplay by YouTuber かんぱり here: https://youtu.be/Px-LtgKICnY?si=9PybDzlp2gXQ38Ls. Note that the user has the background video set to a lower brightness, while my screenshot feature videos set at 100% default brightness.

Figure 6.7: Generalized audio stream and sync train that combines the instrumental and vocal parts into one musical staff.

6.7a



6.7b

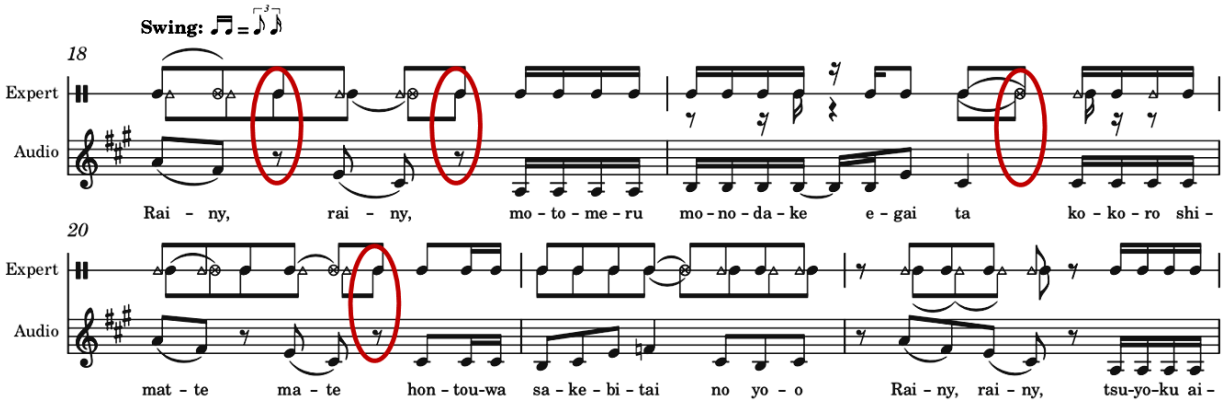


English translation of lyrics:
"I don't have the talent so I suppose I'll just remain here crying all my life
I was envious of the shade of blue in the scene before my eyes"

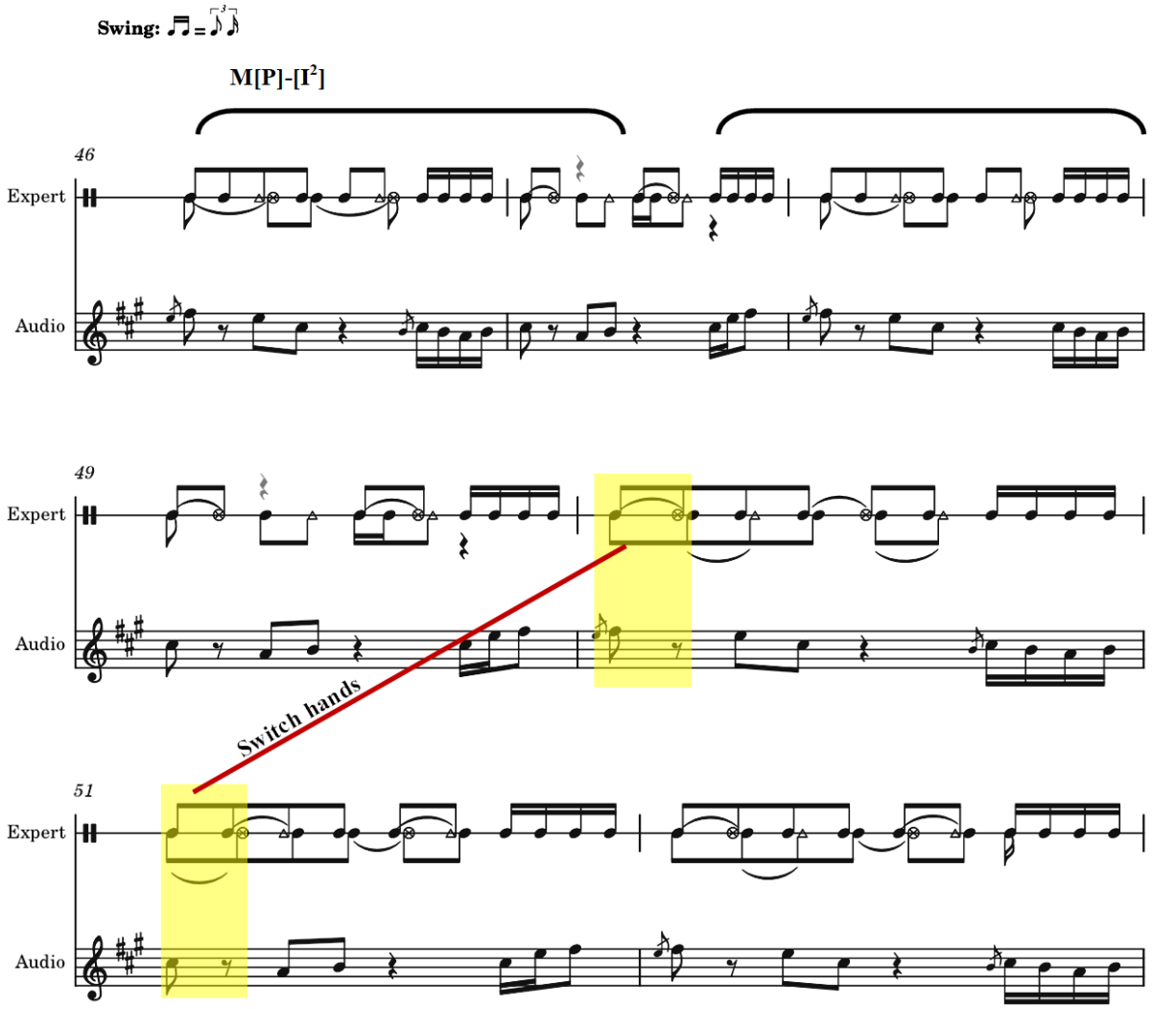
(Project Sekai Wiki, https://projectsekai.fandom.com/wiki/Kagirinaku_Haiiro_e, last accessed January 22, 2024)

Figure 6.8: Generalizations found in the audio stream and beatmap.

6.8a



6.8b. Transcription



6.8c

English translation of lyrics:
“Rainy Rainy, I drew only what I wished to
Closing my heart off - wait! In truth, I want to scream
Rainy Rainy, I wished to be strong
My voice scattered unfeelingly, playing a melody of loneliness”

(Project Sekai Wiki, https://projectsekai.fandom.com/wiki/Kagirinaku_Haiiro_e, last accessed January 22, 2024)

Figure 6.9: Tri-sync identifies synchronous moments across three different streams.

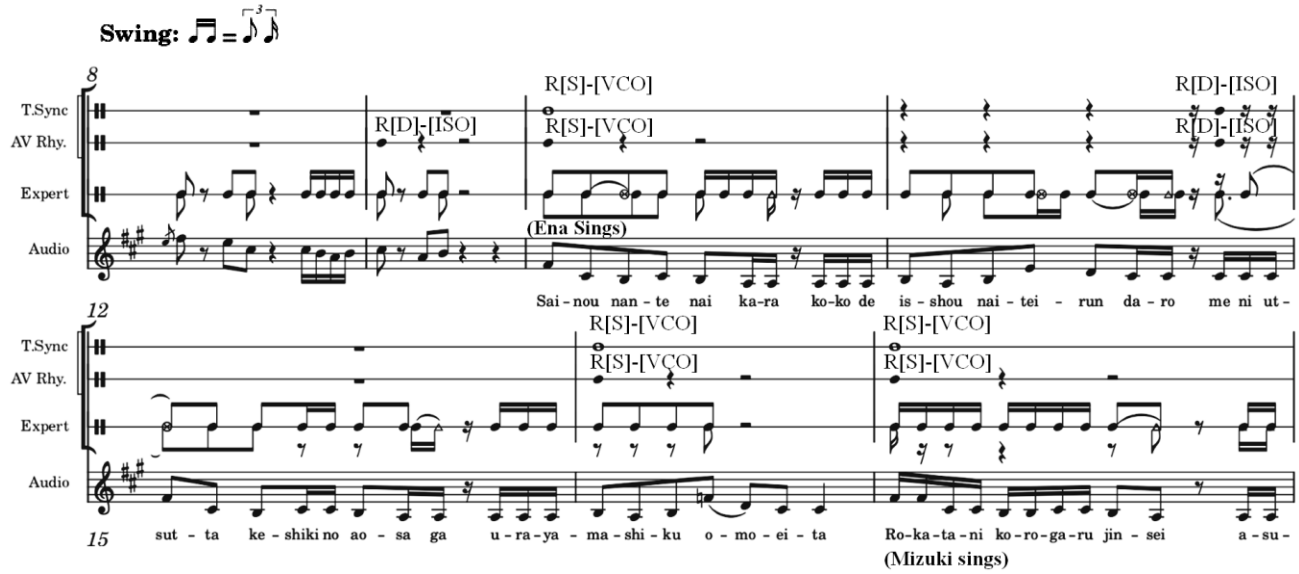
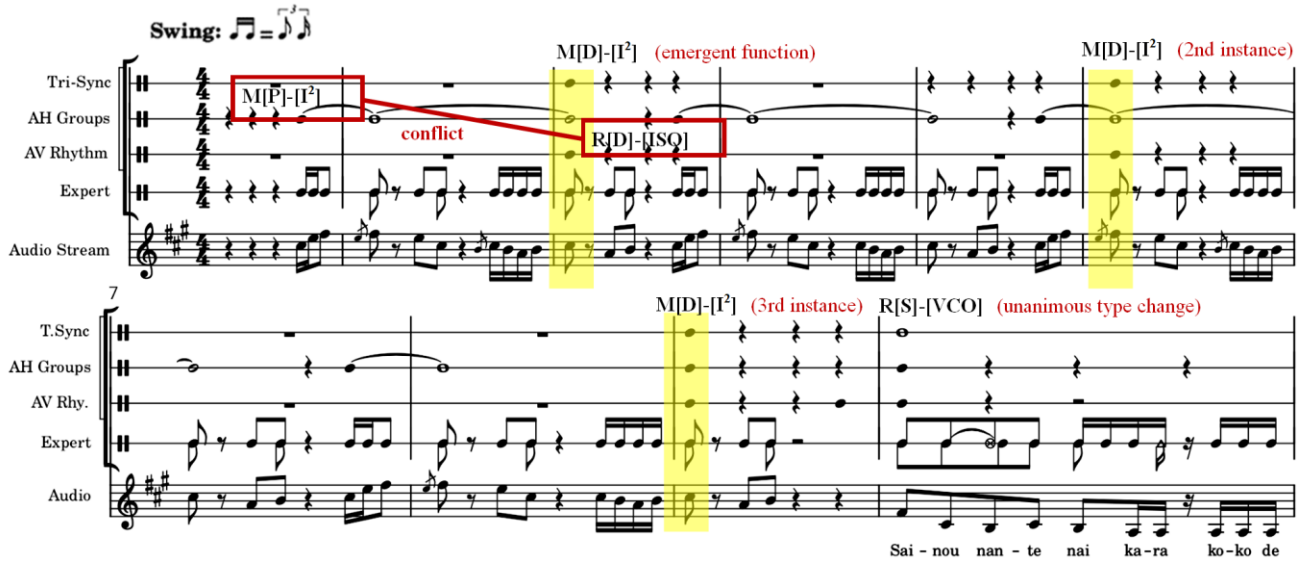


Figure 6.10. Hierarchical sync types in tri-sync. How conflicting multiple lines are solved.

6.10a



6.10b

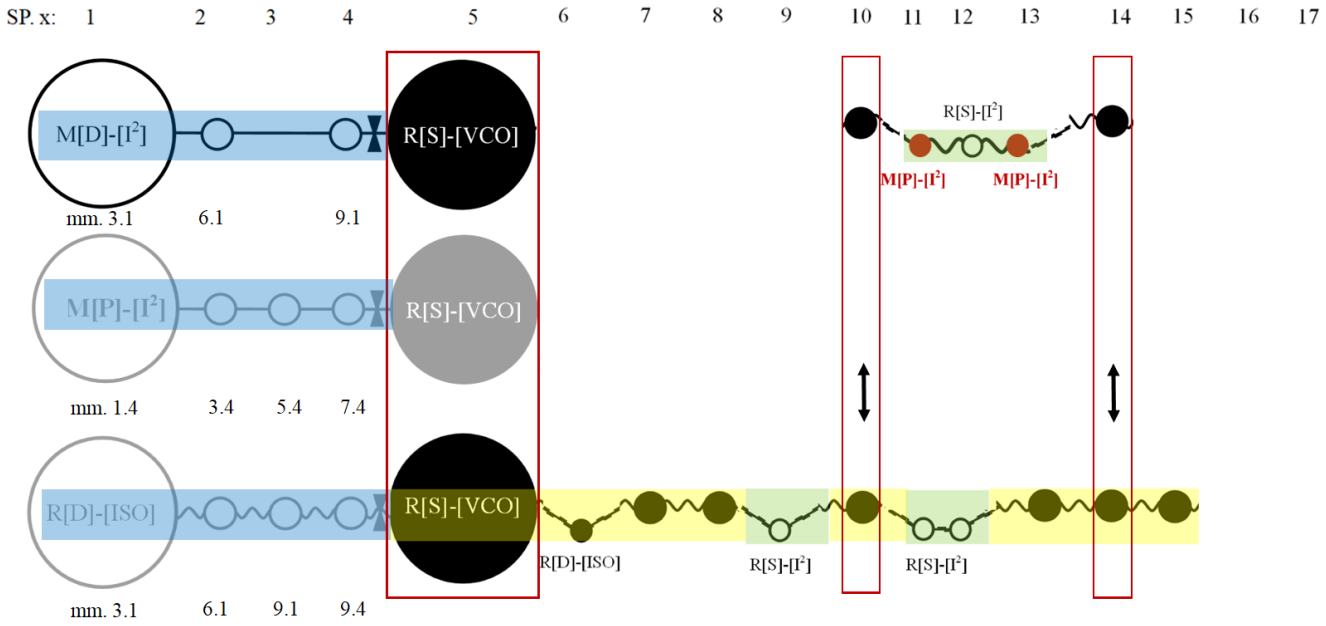
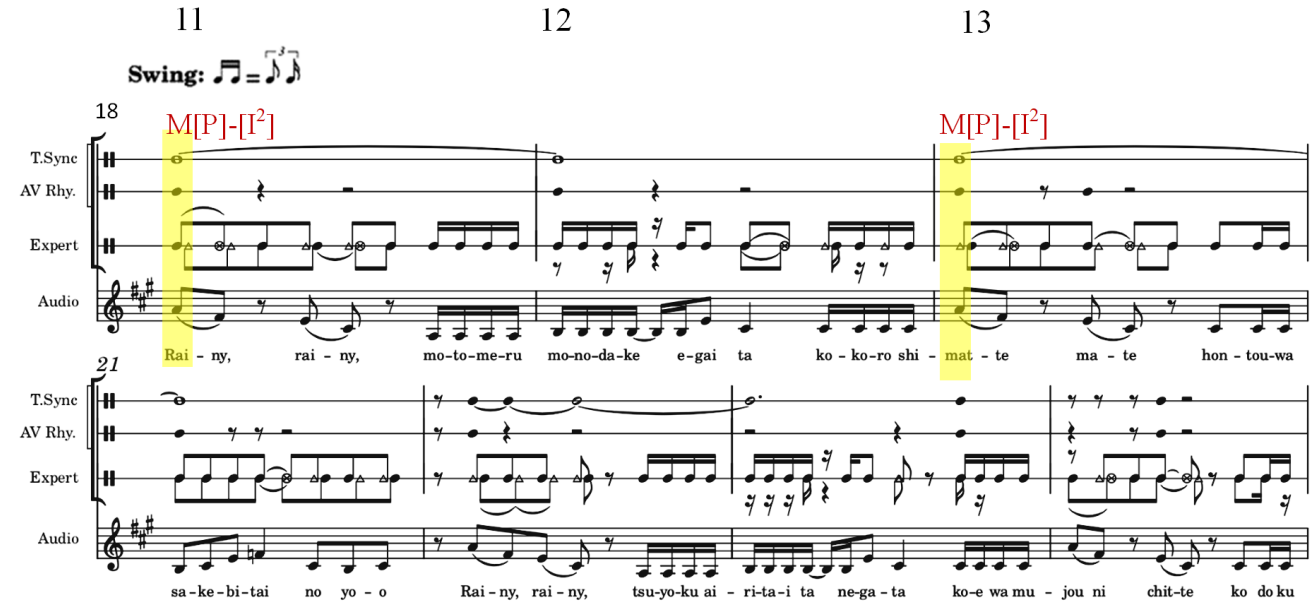
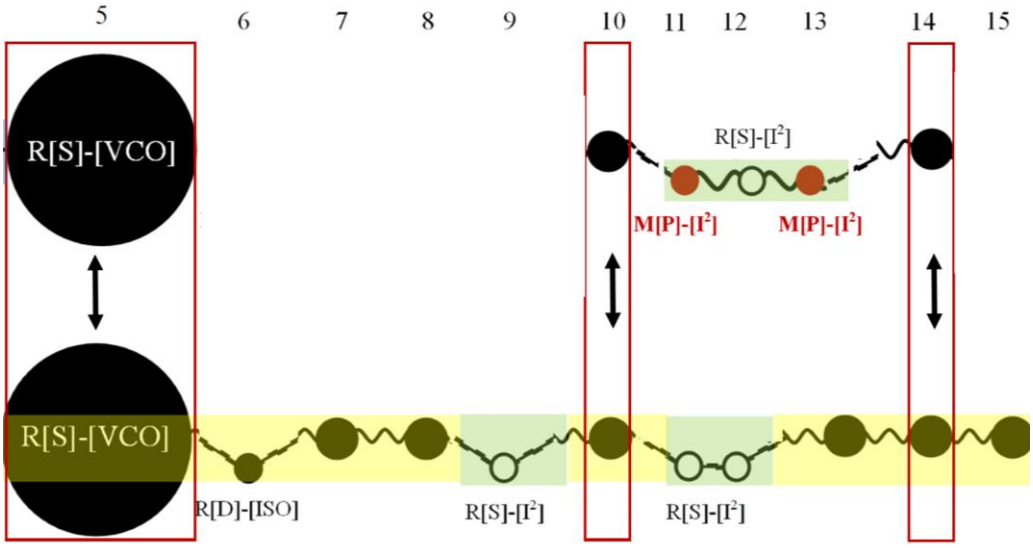


Figure 6.11: Metric overriding at the tri-sync level.[209]

6.11a



6.11b



[209] Gameplay by YouTuber かんぱり here: https://youtu.be/Px-LtgKICnY?si=CRlk_13_Xujt7KYB&t=44. Excerpt starts at 0:44.

Figure 6.12: Color-coded slurs that indicate switching hands to create the same rhythm.[210]

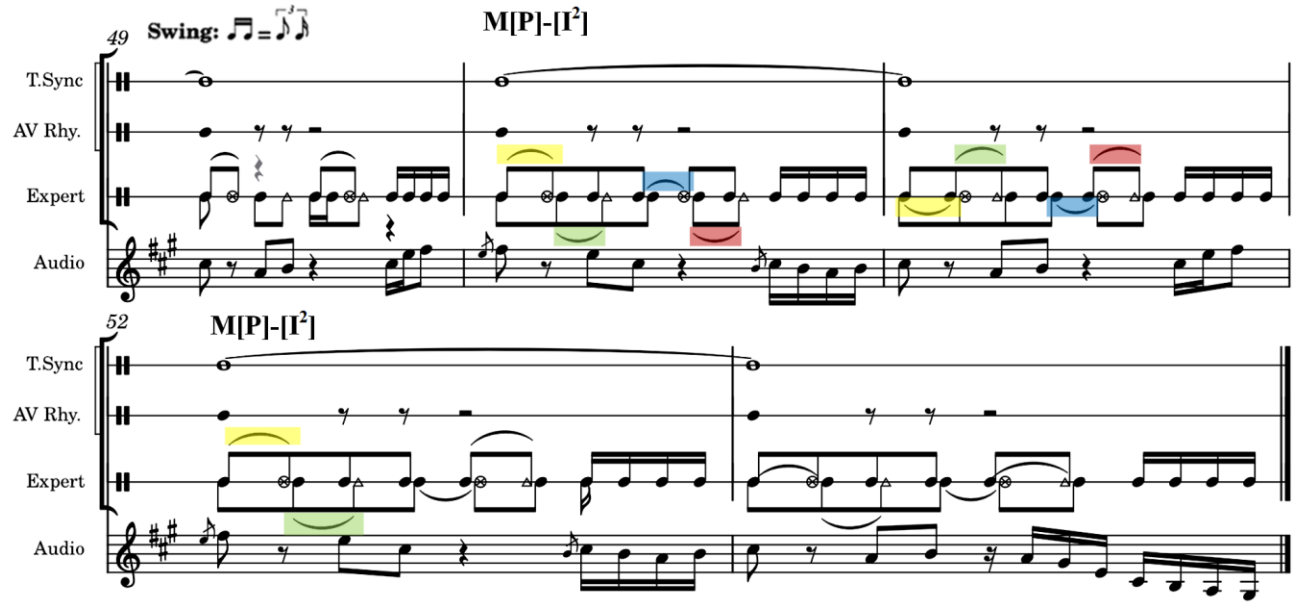
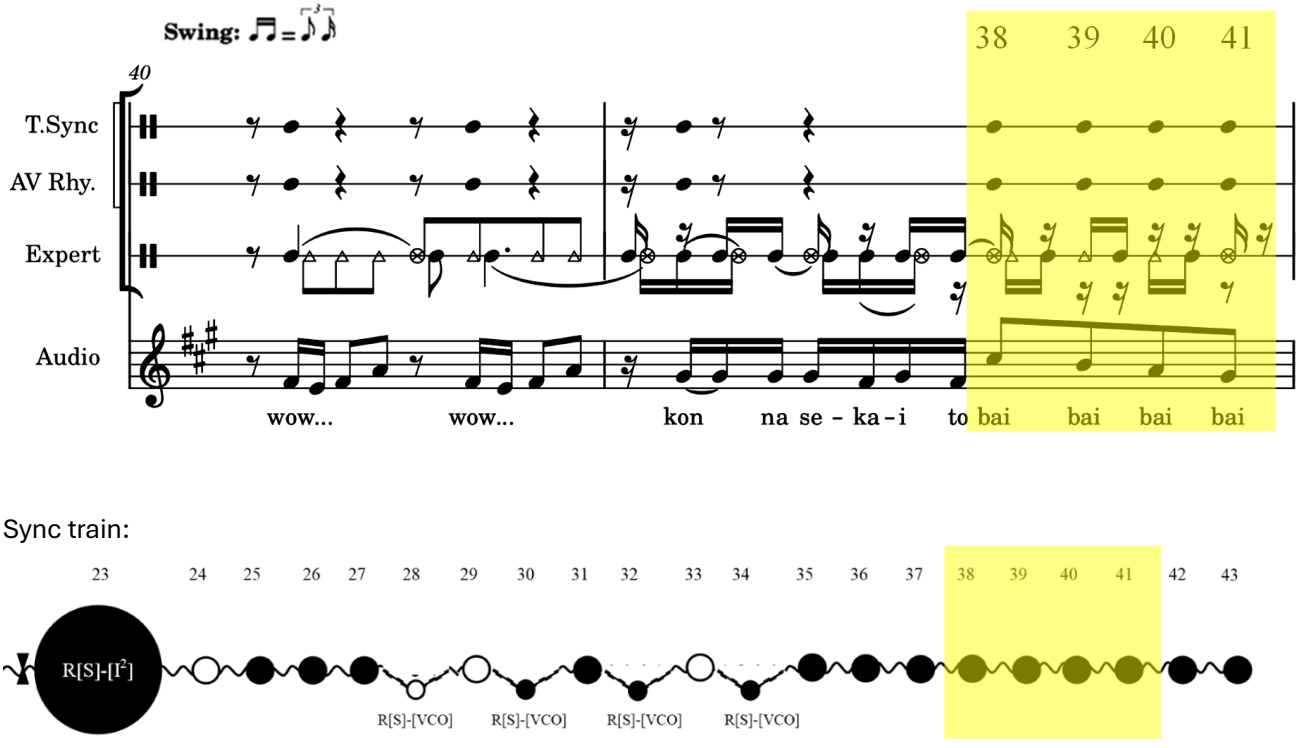
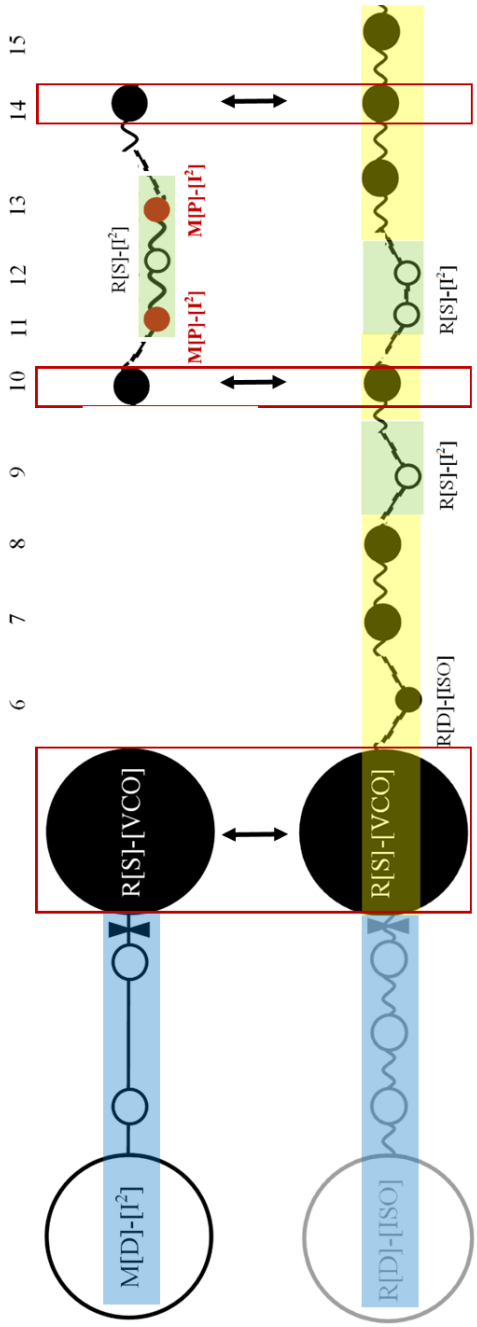


Figure 6.13: Nonmetric instance of tri-sync.



[210] Gameplay by YouTuber かんばり here: https://youtu.be/Px-LtgKICnY?si=WCDbjlgSUr43qs5d&t=114. Excerpt begins at 1:54.

Figure 6.14. Sync train collections switching between tracks.



Tables

Chapter 2 Tables

Table 2.1: List of Rhythmic Sync Types

Music accent	Visual Accent*	Sync Point Type R[x]-[y]	Definition
R[S]-			• Non-periodic sonic events • [S] = [*sonic*] = non-rhythmic emphasis of sound events • Sync to a specific pitch, motif/fragment, timbre, or local harmonic motion; can be musical "sound effects"
R[P]-			• Non-periodic phrasal events • [P] = [*phrasal*] = phrase-structural and formal features • Sync with structural downbeats, cadences, melodies, sentences, periods, formal structures, or functional harmonic progressions
R[D]-			• Non-periodic, temporal durations • [D] = [*duration*] = temporal synchronicities • Sync with rhythmic patterns, velocity (accelerando/ritardando), or any temporal aspects
	[ISO]		• [ISO] = isomorphic • Mirroring of durational structure • Visuals share the same "shape" or "rhythm" as music's temporal structures
		R[D]-[ISO]	Purely rhythmic sync point. Both image and music reflect only temporality and duration. Immediate action is emphasized.
		R[S]-[ISO]	Musical sonic events highlighted by purely rhythmic visual accents.
		~~**R[P]-[ISO]**~~	Null, because [ISO] highlights rhythmic durations over pitch-based functions.
	[ICO]		• [ICO] = iconic • Literal/abstract pictorial representation of non-rhythmic contours • Visuals reflect timbre, pitch, melodies, fragments, motives, etc.
		R[S]-[ICO]	Visuals and music interact non-rhythmically to create a composite meaning.
		R[P]-[ICO]	Visuals connect to a musical phrase or formal unit.
		R[D]-[ICO]	Visuals represent musical durations, without the "mickey-mousing" precision of an R[D]-[ISO] relation. The idea is more important than the action.
	[I^2]		• Simultaneity of isomorphic and iconic interactions. • Visual accents emphasize a musical feature both rhythmically (isomorphic) and non-rhythmically (iconic).
		R[S]-[I^2]	Visuals and music simultaneously create meaning for a sonic event.
		R[P]-[I^2]	Visuals and music simultaneous mirrors a musical phrase; "mickey-mousing"
		R[D]-[I^2]	Visuals and music simultaneously emphasize a musical duration while applying iconic meaning to a timbre or other musical feature.

*Visual accents are defined by action-stop (in-shot), shot-change (between-shot), and camera motion (pan, zoom, etc.) types. The bracketed labels refer to their function in the R[x]-[y] relationship.

Chapter 3 Tables

Table 3.1: [AVC] definition as R[P]-[y] Subset.

Music	Visual*	R[x]-[y]	Subsets	Definition
R[P]-				• Non-periodic phrasal events • [P] = [*phrasal*] = phrase-structural and formal features • Sync with structural downbeats, cadences, melodies, sentences, periods, formal structures, or functional harmonic progressions
	[ICO]			• [ICO] = iconic • Literal/abstract pictorial representation of non-rhythmic features • Visuals reflect timbre, pitch, melodies, fragments, motives, etc.
		R[P]-[ICO]		Visuals connect to a musical phrase or formal unit.
			[AVC]	Part of an [ICO] phrase where cadential notes and harmonies mirror non-rhythmic aspects, as defined by the [ICO] column.
	[I^2]			• Simultaneity of isomorphic and iconic interactions. • Visual accents emphasize a musical feature both rhythmically (isomorphic) and non-rhythmically (iconic).
		R[P]-[I^2]		Visuals and music simultaneous mirrors a musical phrase; "mickey-mousing"
			[AVC]	Part of an [ICO] phrase where cadential notes and harmonies mirror non-rhythmic aspects, as defined by the [ICO] column. However, there is also a rhythmic dimension to the iconic expression of a cadence.

*Visual accents are defined by action-stop (in-shot), shot-change (between-shot), and camera motion (pan, zoom, etc.) types. The bracketed labels refer to their function in the R[x]-[y] relationship.

Table 3.2: [VCO] definition as R[x]-[ICO] Subset.

Music	Visual*	R[x]-[y]	Subsets	Definition
R[S]-				• Non-periodic sonic events • [S] = [*sonic*] = non-rhythmic emphasis of sound events Sync to a specific pitch, motif/fragment, timbre, or local harmonic motion; can be musical "sound effects"
R[P]-				• Non-periodic phrasal events • [P] = [*phrasal*] = phrase-structural and formal features • Sync with structural downbeats, cadences, melodies, sentences, periods, formal structures, or functional harmonic progressions
R[D]-				• Non-periodic, temporal durations • [D] = [*duration*] = temporal synchronicities Sync with rhythmic patterns, velocity (accelerando/ritardando), or any temporal aspects
	[ICO]			• [ICO] = iconic • Literal/abstract pictorial representation of non-rhythmic contours Visuals reflect timbre, pitch, melodies, fragments, motives, etc.
		R[S]-[ICO]		Visuals and music interact non-rhythmically to create a composite meaning.
		R[P]-[ICO]		Visuals connect to a musical phrase or formal unit.
		R[D]-[ICO]		Visuals represent musical durations, without the "mickey-mousing" precision of an R[D]-[ISO] relation. The idea is more important than the action.
			[VCO]	• [VCO] = vococentric • Visuals clarify who/what is singing or sounding
			R[S]-[VCO]	Visuals clarify change of singing voice but does not outline a musical phrase.
			R[P]-[VCO]	Visuals reinforce a singing voice for an entire musical phrase.
			R[D]-[VCO]	Less chance of occurence. Maintains the duration of the voice and may match phrase duration but does not need to adhere to phrase structure.

*Visual accents are defined by action-stop (in-shot), shot-change (between-shot), and camera motion (pan, zoom, etc.) types. The bracketed labels refer to their function in the R[x]-[y] relationship.

Chapter 5 Tables

Table 5.1 Play-by-play for “Close to Gray” by Nightcord at 25:00 (2021).

“Close to Gray” – Nightcord at 25:00 Version				
S.P. (Sync Point)	**Timestamp**	**Sync Type**	**Reroutes**	**Commentary**
1	0:12	R[D]-[ISO]		Camera zooms out from fountain. Is R[D]-[ISO] matches visuals to musical duration.
2	0:19			Characters (5) enter the frame; zooms to long shot of Nightcord group. Rin moves her hand.
3	0:26			Characters initiate motion (head turn).
4	0:27			Characters initiate dance choreography.
5	0:27	R[S]-[VCO]		Cut to close-up shot of Ena. Marks beginning of [VCO]-centricity
6	0:31		R[D]-[ISO]	Reroute. Switch to mid-shot of Kanade and Rin dancing. Their dance is not [ICO] to music or lyrics.
7	0:34			Cut to close-up of Ena.
8	0:36			Cut to wide shot of Mizuki; [VCO] focus.
9	0:38		R[S]-[I^2]	Reroute. R[S]-[I^2] highlights the choreography, expressing sounds and lyrics (“jinsei”/“asphalt”).
10	0:41			Cut to mid shot of Mizuki. Editing constantly refocuses [VCO].
11	0:45		R[S]-[I^2]	Cut to wide shot of group. R[S]-[I^2] Matches “rainy, rainy."
12	0:48		R[S]-[I^2]	Cut to camera following Ena.
13	0:50			Cut; [VCO] focus on singing ensemble.
14	0:51			Cut; [VCO] focus on Ena.
15	0:52			Cut; close-up shot of Ena.
16	0:54			Cut; [VCO] to Kanade and Mizuki. Rin (also singing) is not in the shot.
17	0:58			Cut; [VCO] to Kanade and Mafuyu.

18	1:00		R[S]-[I^2]	Reroute. Close-up of Ena *expressing* (not singing) Kanade and Mafuyu's lines.
19	1:03, and then leading into Sp. 22…			Cut; [VCO] to Rin, whose voice was disembodied for brief seconds.
20			R[S]-[I^2]	Camera follows Rin's choreography.
21			R[S]-[I^2]	
22	1:09			Cut; closeup to Ena singing "mitomete wa kuranai no?" (trans. "Won't you recognize me?"[211]
23	1:12	R[S]-[I^2]		Cut; wide shot of group choreography, which takes priority in screentime through R[S]-[I^2] sync.
24	1:14			In-shot choreography highlights chord roll.
25	1:15			Cut to two characters but adds no meaning other than durational matching.
26	1:17			
27	1:20			Not [VCO] because visuals do not focus on a singular voice.
28	1:21		R[S]-[VCO]	Cut to Rin. [VCO] focus on Rin.
29	1:23			Character Rin spins with chord roll.
30	1:23		R[S]-[VCO]	Cut to Mafuyu in motion.
31	1:25			Cut to wide shot of choreography.
32	1:27		R[S]-[VCO]	Cut to Ena, [VCO] focus.
33	1:30			Zoom out to wide shot of choreography. Synced with sung "wow, wow…"
34	1:32		R[S]-[VCO]	Cut to Mizuki, [VCO] focus.
35	1:34			Cuts made to dancing characters, but fidelity is not prioritized.
36	1:35			
37	1:37			Cut to Ena synced to harmonic changes and lyrics.
38	1:37			Cuts (rapid) synced to "bye, bye, bye, bye…" in the lyrics. Compounded with in-shot dance motions.
39	1:37			
40	1:37			
41	1:37			

[211] Lyrics translated and provided by Vocaloidlyrics.fandom.com: https://vocaloidlyrics.fandom.com/wiki/%E9%99%90%E3%82%8A%E3%81%AA%E3%81%8F%E7%81%B0%E8%89%B2%E3%81%B8_(Kagirinaku_Haiiro_e). Last accessed September 2023.

42	1:39			Cuts synced to “wow, wow…” All syncs are dance focused. Cuts to Mizuki and Mafuyu (1:39) Cuts to Kanade and Rin (1:40) Cuts to Rin (1:42)
43	1:40			
44	1:42			
45	1:43			Cut, then zoom to Ena.
46	1:47			Ena pushes camera upwards.
47	1:48	M[P]-[I^2]		Flowers frame the addition of color.
48	1:50			Cuts sync to the instrumental track’s downbeats at the beginning of every musical measure. More stable without lyrics and vocal melody. A metric collection is formed.
49	1:52			
50	1:54			
51	1:56			
52	1:59			
53	2:01			
54	2:03			
55	2:05			
56 (end)	2:10			Fade out to white screen.

Table 5.2: Play by play for “Close to Gray” by Surii (2021).

“Close to Gray” – Three and Len Version				
S.P. (Sync Point)	**Timestamp**	**Sync Type**	**Reroutes**	**Commentary**
1	0:00	R[D]-[ISO]		Fade into Rainy City; pan downwards.
2	0:10			Cut to sleeping white bird.
3	0:12			Fade in title card “限りなく灰色へ.”
4	0:19			Cut to main character (MC) open eyes.
5	0:20			Cut to MC sitting up from bed.
6	0:23			Zoom out for wide shot.
7	0:26			Cut to matchbox.
8	0:27			Cut to submerged feet.
9	0:34	R[D]-[I^2]		Cut syncs to “denpa” (digital/electric) and “sekai” (world).
10	0:35			
11	0:36			Swipe transitions sync shots of midi keyboard to “Rainy, rainy…”
12	0:37			
13	0:38		R[D]-[ISO]	Cut to MC working at computer.
14	0:41		R[D]-[ISO]	Cut to mouse moving across file explorer.
15	0:43		R[D]-[ICO]	Cut to iconic visualization of MC’s frustrations. Reroute, but does not match prior R[D]-[ISO]. Tripartite preference not met, and R[D]-[I^2].
16	0:46			Swipe transitions sync shots of midi keyboard to “Rainy, rainy…”
17	0:46			
18	0:47	R[D]-[ISO]		Cut: R[D]-[ISO] return to MC looking at computer.
19	0:49			Cut to file explorer. In shot motion does not match to an [S] or [P] accent.
20	0:52			Cut back to MC, close-up shot.
21	0:53		R[D]-[I^2]	Cut to raven and in-shot effect.
22	0:54			Cut to MC’s “other self.”
23	1:01			Cut to “other self” and raven. Raven flies of and “other self” is masked.
24	1:02		R[S]-[ICO]	Distorted glissando synced to fizzled black screen.
25	1:03			Cut to MC at street junction.
26	1:07		R[D]-[I^2]	Cut to close-up of MC mouthing lyrics.
27	1:08	R[S]-[I^2]		Cuts sync to harmonic changes in the music.
28	1:08			

29	1:10			
30	1:11			
31	1:12	R[D]-[ISO]		Cut to and zoom out from MC.
32	1:14			Cut over-shoulder to MC watching Nightcord at 25:00 version.
33	1:18			Cut to matchbox.
34	1:19			MC dissipates.
35	1:21			Cut; MC appears in an empty world.
36	1:25			Character (bird) enters.
37	1:25			MC turns.
38	1:25			Cut synced to “wow, wow…” Zoom out to wide shot.
39	1:29			Cut to MC; Bird points leftward.
40	1:30			Cut to hooded “other self.”
41	1:34			Cut to MC.
42	1:36			Cut to “other self.”
43	1:37			Other self “fires.”
44	1:38			Screen shatters.
45	1:39	R[S]-[I^2]		Cut to MC waking up.
46	1:39			Swipe transition between keyboard and MC.
47	1:40			
48	1:40		R[D]-[ISO]	MC sits up.
49	1:41			Shots switch between MC outside of and inside of room.
50	1:42			
51	1:44		R[D]-[ISO]	Cut to matchbox.
52	1:45		R[D]-[ISO]	Hand closes on matchbox.
53	1:47			Cut to city building.
54	1:48			Swipe transition to “other self.”
55	1:51			In-shot fizzle effect.
56	1:51			Cut to raven.
57	1:52		R[D]-[ISO]	Cut to white bird.
58				In-shot effect.
(End of first half)				

Glossary of Analytical Terms

Audiovisual Cadence: visually reinforced "ends" and "closures" of musical phrases, section, or formal structures found in film music sequences.

Aurality: Contextual spaces that feature more non-coincidental musical [x] than visual markedness.

Beatmap: And aggregate term for fixed rhythmic patterns in a rhythm game. These patterns are animated, visible, and tactile.

Function: A set of conditions and expectations formed by recurrence or non-recurrence of emergent meanings. Detailed in Chapter 2.

Haptic Sync: Synchronicity between an audience-turned-player and a media device that hosts a musical media experience.

Meaning: Dialectic rhythms and hermeneutic narratives created by audiovisual interactions (i.e., sync points). Emergent meanings form *expectations* for recurring patterns of synchronicity.

Reroutes: Parts of a sync train that wedge out from the graph to indicate the presence of a secondary type of sync point relation.

Synchronicity: The action of sound and image rhythmically mirroring each other in fluctuating periodicities. Detailed in Chapter 2.

Sync Collections: groups of recurring sync points of the same types.

Sync Points: Precise moments when sound and image rhythmically mirror each other in fluctuating periodicities. They are *conditions* for function, and their functions are defined by the n[x]-[y] formulaic label. Detailed in Chapters 2, 3, and 4.

Sync Train: Larger groups of *same* and *different* sync collections that also contextualize what happens *between* sync points.

Totality: Contextual spaces that feature equal *non-coincidental* musical [x] and visual [y] markedness.

Visuality: Contextual spaces that feature more *non-coincidental* visual [y] than musical markedness.

Vococentric Shot Change: Synchronous visual accents that emphasize the singing voice, with fidelity to the singing voice (showing who is singing at the moment).

Bibliography

Afra, Kia. "'Vertical Montage' and Synaesthesia: Movement, Inner Synchronicity, and Music–Image Correlation in *Alexander Nevsky* (1938)." In *Music, Sound and the Moving Image* 9, no. 1 (2015): 33-61. doi: 10.3828/msmi.2015.2.

Altman, Rick. "Visual Representation of Film Sound as an Analytical Tool." In *The Oxford Handbook of Film Music Studies*. Edited by David Neumeyer. Oxford: Oxford University Press, 2013.

_____. "General Introduction: Cinema as Event." In *Sound Theory, Sound Practice*. Edited by Rick Altman, 1–14. New York: Routledge, 1992.

_____. "Introduction: Sound/History," in *Sound Theory, Sound Practice*. Edited by Rick Altman, 113-125. New York: Routledge, 1992.

_____. "Sound Space." In *Sound Theory/Sound Practice*. Edited by Rick Altman, 46–64. New York: Routledge, 1992.

Auerbach, Brent. "Pedagogical Applications of the Video Game Dance Dance Revolution to Aural Skills Instruction." In *Music Theory Online* 16, 1 (2010). https://mtosmt.org/issues/mto.10.16.1/mto.10.16.1.auerbach.html.

Bartig, Kevin. *Sergei Prokofiev's Alexander Nevsky.* Oxford: Oxford University Press, 2017.

Bordwell, David and Kristin Thompson. *Film art: an introduction*. Boston: McGraw-Hill, 2004.

Bradshaw, Michael. "Prokofiev's New Simplicity and Expressive Form." Phd. dissertation. University of Queensland, 2018.

Bribitzer-Stull, Matthew. *Understanding the Leitmotif: From Wagner to Hollywood Film Music*. Cambridge: Cambridge University Press, 2015.

Buhler, James. *Theories of the Soundtrack*. New York: Oxford University Press, 2019.

Caplin, William E. "Theories of musical rhythm in the eighteenth and nineteenth centuries." In *The Cambridge History of Western Music Theory*. Ed. Thomas Christensen. United Kingdom: Cambridge University Press, 2002.

_____. *Analyzing Classical Form: An approach for the classroom*. New York: Oxford University Press, 2013.

Cochran, Alfred. "Style, Structure, and Tonal Organization in the Early Film Scores of Aaron Copland." Ph.D. dissertation. The Catholic University of America, 1986.

Collins, Karen, "Interacting with Sound." In *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge Massachusetts: MIT Press, 2013.

_____. "Embodying Game Sound in Performance: The Real and the Virtual." In *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge, Massachusetts: MIT Press, 2013.

Curtis, Scott. “The Sound of the Early Warner Bros. Cartoons.” In *Sound Theory, Sound Practice*. Edited by Rick Altman. New York: Routledge, 1992.

Davies, David. “Animation.” In *The Palgrave Handbook of the Philosophy of Film and Motion Pictures*. Edited by Carroll Noel, Laura T. Di Summa, and Shawn Loht. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-19601-1_8.

Deleuze, Gilles. *Cinema 1: The Movement-Image.* Minneapolis: University of Minnesota Press, 1986.
_____. *Cinema 2: The Time-Image.* Minneapolis: University of Minnesota Press, 1989.

Donnelly, Kevin J. “The Triple Lock of Synchronization.” In *The Cambridge Companion to Video Game Music*. Edited by Melanie Fritsch and Tim Summers. Cambridge: Cambridge University Press, 2021. doi:10.1017/9781108670289.008.

Eisenstein, Sergei. *Film Form: Essays in Film Theory*. Translated by Jay Leyda. New York: Harcourt, Inc., 1977.
_____. *Film Sense*. Translated by Jay Leyda. New York: Harcourt, Inc., 1943.
_____. *On Disney*. Translated by Jay Leyda. London: Methuen, 1988.

Ermolaeva, Katya. “Audiovisual Montage in Ivan the Terrible.” In *Rethinking Prokofiev*. Edited by Christina Guillaumier and Rita McAllister. New York City: Oxford University Press, 2020. doi:10.1093/oso/9780190670764.001.0001.

Flaxman, Gregory. "Once More, with Feeling: Cinema and Cinesthesia." In *SubStance* 45, 3: 174-189 (2016). muse.jhu.edu/article/634224.

Ford, Fiona. “The Film Music of Edmund Meisel (1894–1930).” PhD dissertation. University of Nottingham, 2011. http://eprints.nottingham.ac.uk/id/eprint/12271.

Goldmark, Daniel. *Tunes for ‘Toons: Music and the Hollywood Cartoon*. Los Angeles: University of California Press, 2005.

Grasso, Julianne. “Video Game Music, Meaning, and the Possibilities of Play.” PhD diss., University of Chicago, 2020.

Hasty, Christopher. *Meter as Rhythm*. New York: Oxford University Press, 1997.

Hauptmann, Moritz. *The Nature of Harmony and Metre*. Trans. W. E. Heathcote. London: S. Sonnenschein, 1888.

Jacobs, Lea. *Film Rhythm after Sound: Technology, Music, and Performance*. Oakland: University of California Press, 2015.
_____. "A Lesson with Eisenstein: Rhythm and Pacing in Ivan the Terrible, Part I." *Music and the Moving Image* 5, no. 1 (2012): 24-46. doi:10.5406/musimoviimag.5.1.0024.

Jarvis, Brian Edward and John Peterson. “Alternative Paths, Phrase Expansion, and the Music of Felix Mendelssohn.” In *Music Theory Spectrum* 41, No. 2 (2019): 187–217. https://doi.org/10.1093/mts/mtz009.

Kaae, Jesper. “Theoretical Approaches to Composing Dynamic Music for Video Games.” In *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*. Edited by Karen Collins. Aldershot: Ashgate, 2008.

Kane, Brian. “Excavating Lewin’s ‘Phenomenology.” In *Music Theory Spectrum* 33, no. 1 (2011): 27–36. https://doi.org/10.1525/mts.2011.33.1.27.

Karpinski, Gary. *Aural Skills Acquisition: The development of listening, reading, and performing skills in college-level musicians*. New York: Oxford University Press, 2000.

Kellman, Noah. “Designing Interaction between the Player and the Music.” In *The Game Music Handbook: A Practical Guide to Crafting an Unforgettable Musical Soundscape*. New York, 2020; online ed., Oxford Academic, Sep. 23, 2021. https://doi.org/10.1093/oso/9780190938680.001.0001, accessed 21 Aug. 2023.

Kepley, Vance. “Pudovkin, Socialist Realsim, and the Classical Hollywood Style.” In the *Journal of Film and Video* 47, no. 4 (1995): 3-16. http://www.jstor.org/stable/20688078.

Kulezic-Wilson, Danijela. *The Musicality of Narrative Film*. Basingstoke: Palgrave Macmillan, 2015.

Langkjær, Birger. "Audiovisual Styling and the Film Experience: Prospects for Textual Analysis and Experimental Approaches to Understand the Perception of Sound and Music in Movies." In *Music and the Moving Image* 8, no. 2 (2015): 35–47. doi:10.5406/musimoviimag.8.2.0035.

Lehman, Frank. *Hollywood Harmony*: *Musical Wonder and the Sound of Cinema*. New York: Oxford University Press, 2018.

_____. “Hollywood Cadences: Music and the Structure of Cinematic Expectation.” In *Music Theory Online* 19, no. 4 (2013). doi: 10.30535/mto.19.4.2

_____. “Methods and Challenges of Analyzing Screen Media.” In *The Routledge Companion to Screen Music and Sound*. Edited by Miguel Merra, Ron Sadoff, and Ben Winters: 497–516. New York: Routledge, 2017b.

Lerdahl, Fred and Ray S. Jackendoff. *A Generative Theory of Tonal Music*. Cambridge: MIT Press, 1983.

_____. “On the Theory of Grouping and Meter,” *The Musical Quarterly* 67, no. 4 (1981): 479–506, http://www.jstor.org/stable/742075.

Leslie, Esther. *Hollywood Flatlands: Animation, Critical Theory, and the Avant-Garde*. London: Verso, 2002.

Lewin, David. *Generalized Musical Intervals and Transformations*. New Haven: Yale University Press, 1987.
_____. “Music Theory, Phenomenology, and Modes of Perception.” In *Music Perception: An Interdisciplinary Journal* 3, no. 4 (1986): 327–92. https://doi.org/10.2307/40285344.
_____. With David Bard-Schwarz and Richard Lawrence Cohn. “Morgengruß by David Lewin.” In *David Lewin's Morgengruß: Text, Context, Commentary*. New York: Oxford University Press, 2015. Edited by David Bard-Schwarz and Richard Cohn. doi: https://doi-org.uc.idm.oclc.org/10.1093/acprof:oso/9780199844784.003.0002

London, Justin. *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press, 2012.

McAdams, Stephen and Albert Bregman. “Hearing Musical Streams.” In *Computer Music Journal* 3, 4: 26–60 (1979). http://www.jstor.org/stable/4617866.

Manvell, Roger and John Huntley. *The Technique of Film Music*. New York: Hastings House, 1957.

Miler, Kiri. *Playing Along: Digital Games YouTube and Virtual Performance*. Oxford: Oxford University Press, 2012.

Mirka, Danuta. “Musical Meter between Composition and Perception.” In *Metric Manipulations in Haydn and Mozart*. New York: Oxford University Press, 2009.

Monaco, James. “Film Theory: Form and Function.” In *How to Read a Film*. New York: Oxford University Press, 2000.

Motazedian, Tahirih. "To Key Or Not to Key: Tonal Design in Film Music." Ph.D. dissertation. Yale University, 2016.

Murphy, Scott. “Transformational Theory and the Analysis of Film Music.” In *The Oxford Handbook of Film Music Studies*. Edited by David Neumeyer. Oxford: Oxford University Press, 2013. doi:10.1093/oxfordhb/9780195328493.013.019.
_____. “The Major Tritone Progression in Recent Hollywood Science Fiction Films.” Music Theory Online 12/2 (2006).

Nattiez, Jean-Jacques. *Musical Analyses and Musical Exegesis: The Shepherd's Melody in Richard Wagner's Tristan and Isolde*. Translated by Joan Campbell Huguet. New York: University of Rochester Press, 2021. doi:10.1017/9781800103634.

Neumeyer, David. *Meaning and Interpretation of Music in Cinema*. Bloomington: Indiana University Press, 2015.
_____. "Film Music Analysis and Pedagogy." In *Indiana Theory Review* 11 (1990): 1-27. http://www.jstor.org/stable/24045976.
_____. “Tonal Design and Narrative in Film Music: Bernard Herrmann’s ‘A Portrait Of Hitch’ and ‘The Trouble With Harry.’” in *Indiana Theory Review* 19 (1998): 87-123.

O'Meara, Daniel. "Rocksmith and the Shaping of Player Experience." In *Music Video Games: Performance, Politics, and Play*. New York: Bloomsbury Academic, 2016.

Ōtsuka, Eiji, and Thomas Lamarre. "An Unholy Alliance of Eisenstein and Disney: The Fascist Origins of Otaku Culture." In *Mechademia* 8 (2013): 251-277. doi:10.1353/mec.2013.0002.

Pearlman, Karen. "On Rhythm in Film Editing." In *The Palgrave Handbook of the Philosophy of Film and Motion Pictures*. Edited by Carroll Noel, Laura T. Di Summa, and Shawn Loht. Cham: Springer International Publishing, 2019. doi: https://doi.org/10.1007/978-3-030-19601-1_7.

Prendergast, Roy. *Film Music: A Neglected Art*. New York: W.W. Norton, 1992.

Pudovkin, Vsevolod. *Film Technique and Film Acting*. New York: Lear, 1949.

Rahn, Steven. "Motivic Transformation in the Dark Knight Trilogy (2005–2012)." *In Music and the Moving Image* 12, no. 2 (2019): 40-60. www.jstor.org/stable/10.5406/musimoviimag.12.2.0040.

Reyland, Nicholas. "Corporate Classicism and the Metaphysical Style: Affects, Effects, and Contexts of Two Recent Trends in Screen Scoring." In *Music, Sound, and the Moving Image* 9, no. 2 (2016): 115-130.

Richards, Mark. "Tonal Ambiguity in Popular Music's Axis Progressions." In *Music Theory Online* 23, 3 (2017). doi: 10.30535/mto.23.3.6.

Rodman, Ronald. ""There's No Place Like Home": Tonal Closure and Design in The Wizard of Oz." In Indiana Theory Review 19 (1998): 125-43. http://www.jstor.org/stable/24044541.

_____. "Tonal Design and the Aesthetic of Pastiche in Herbert Stothart's Maytime." In *Music and Cinema*. Edited by James Buhler, Caryl Finn, and David Neumeyer: 187–206. Hanover: Wesleyan University Press.

Rothstein, William N. *Phrase Rhythm in Tonal Music.* New York: Schirmer Books, 1989.

Schneller, Tom. "Easy to Cut: Modular Form in the Film Scores of Bernard Herrmann." In *Journal of Film Music* 5 (2013): 127-151.

Shultz, Peter. "Rhythm Sense: Modality and Enactive Perception in Rhythm Heaven." In *Music Video Games: Performance, Politics, and Play*. New York: Bloomsbury Academic, 2016.

Smith, Susan. "The Animated Film Musical." In *The Oxford Handbook of the American Musical*. Edited by Raymond Knapp, Mitchell Morris, and Stacy Wolf, 167-178. New York: Oxford University Press, 2011.

Sobchak, Vivian. “What My Fingers Knew: The Cinesthetic Subject, or Vision in the Flesh.” In *Carnal Thoughts: Embodiment and Moving Image Culture*. Berkeley: University of California Press, 2004.

Sologub, Olga. “Sergei Prokofiev’s Piano Sonata No. 8, op. 84 and Symphony No. 5, op. 100: Neo-Riemannian and Kholopovian Perspectives.” PhD diss. University of Manchester, 2014.

Stilwell, Robynn. "Music in Films: A Critical Review of Literature, 1980-1996." In *the Journal of Film Music* 1, no. 1 (2009): 19–61. https://doi-org.uc.idm.oclc.org/10.1558/jfm.v1i1.19.

Strand, Joachim. “The Cinesthetic Montage of Music-Video: Hearing the Image and Seeing the Sound.” MCA Thesis. Curtin University of Technology, 2006.

Thomas, Frank, and Ollie Johnston. *The Illusion of Life: Disney Animation*. New York: Hyperion, 1995.

Wells, Paul. *Understanding Animation*, New York: Routledge, 1998.
_____. *Animation and America*. New Brunswick: Rutgers University Press, 2002.

Wiseman, Frederick. "A Symposium on Editing: On Editing." In *The Threepenny Review*, no. 113 (2008): 16-19.

Yu, Cheuk Ling. “Japanese Anime Idol Music in the Media Mix: A Case Study of the Love Live Idols.” Master’s thesis. University of San Diego, 2021. Retrieved from ProQuest, https://uc.idm.oclc.org/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fdissertations-theses%2Fjapanese-anime-idol-music-media-mix-case-study-em%2Fdocview%2F2625044103%2Fse-2%3Faccountid%3D2909.

Media References

Disney. “Hercules | Zero to Hero | Lyric Video | Disney Sing Along.” YouTube video, 2:16. https://youtu.be/yOL-EJZjmp0?si=5zaxHmNKM-mDtPQT.

Disney. “The Little Mermaid | Poor Unfortunate Souls | Disney Sing-Along.” YouTube video, 4:11. https://youtu.be/Gi58pN8W3hY?si=5jpwMKmh2flgzruR.

DisneyMusicVEVO. “Kristen Bell, Santino Fontana - Love Is an Open Door (From "Frozen"/Sing-Along).” YouTube video, 2:05. https://youtu.be/kQDw88hEr2c?si=V-ixJYjIKXOb64Sf

DisneyMusicVEVO. “The Little Mermaid - Under the Sea (from The Little Mermaid) (Official Video).” YouTube video, 3:01. https://youtu.be/GC_mV1IpjWA?si=fgue5Wp2HbEi5RKl.

DisneyMusicVEVO. “Where You Are (From “Moana”/Sing-Along).” YouTube video, 3:16. https://youtu.be/RTWhvp_OD6s?si=bhUeaIlXKei49Ll0.

Fix Me a Scene. “Alexander Nevsky (1938) — Battle On The Ice.” YouTube video. https://youtu.be/IcPixaWL2Pg?si=jhsI-eePg7_hj8YS

HATSUNE MIKU: COLORFUL STAGE!. “HATSUNE MIKU: COLORFUL STAGE! - Close to gray by Three 3D Music Video - Nightcord at 25:00.” YouTube video. https://youtu.be/3054q5rA5JQ?si=OgQQa3gLe1JeN7tS

Kinema Citrus. 少女☆歌劇 レヴュースタァライト (*Shōjo Kageki Revyū Sutāraito*). Tomohiro Furukawa, director. Sentai Filmworks, 2018, streaming video, 25:26. Accessed June 2024. HiDIVE. https://www.hidive.com/season/20527.

Mosfilm. “Alexander Nevsky | DRAMA | FULL MOVIE | by Sergei Eisenstein. YouTube video, 55:26. https://youtu.be/Gq4PaJfod4w?si=oUU2B9FokDotRiW6

Prokofiev, Sergey. Alexander Nevsky: Cantata for Mezzo-Soprano Solo, Mixed Chorus and Orchestra, Op. 78. Tokyo: Zen-On Music Co., 1999.

Russia. “Russia RU - 📽️ #OTD in 1938 Sergey Eisenstein's historical epic "Alexander Nevsky" premiered - to astounding success. It was a visual feast of unprecedented scale (battles!) with an OST by Sergey Prokofiev himself! 💬 Quote: "Whoever will come to us with a sword, from a sword will perish.” Internet Archive video, 0:25. https://archive.org/details/twitter-1466145850166587401

Saint-Saëns, Camille. *Le Carnaval des Animaux*. Edited by Auguste Durand and Fils. Paris:

Durand, 1922. Accessed June 2024. https://vmirror.imslp.org/files/imglnks/usimg/0/06/IMSLP576538-PMLP6099-Le_carnaval_des_animaux_-...-Saint-Sae-ns.pdf.

Saint-Saëns, Camille. *Le Carnaval des Animaux*. Arranged for piano, 4 hands by Lucien Garban. Paris: Durand & Cie., 1922. Reprint, Boca Raton: Masters Music, n.d. (1998). Accessed June 2024. https://ks15.imslp.org/files/imglnks/usimg/1/10/IMSLP13513-Saint-Saens_Carnaval_des_Animaux_piano_4_hands.pdf.

Surii (すりぃ). 限りなく灰色へ (*Close to Gray*). Arranged by ShiA. In *Musescore.com*. Accessed June 2024. https://musescore.com/user/34905762/scores/6656120.

Walt Disney Animation Studios. “Walt Disney Animation Studios' Steamboat Willie.” YouTube video, 5:27. https://youtu.be/BBgghnQF6E4?si=zkXbYvViRywtL-F8.

Walt Disney Feature Animation. *Mulan*. Barry Cook and Tony Bancroft, directors. Walt Disney Studios, 1998, streaming video, 1:36:00. Accessed June 2024. Disney Plus. https://www.disneyplus.com/movies/mulan/85wmj4hahA0B

Walt Disney Feature Animation. *The Hunchback of Notre Dame*. Gary Trousdale and Kirk Wise, directors. Walt Disney Studios, 1996, streaming video, 1:37:00. Accessed July 2024. Disney Plus. https://www.disneyplus.com/movies/the-hunchback-of-notre-dame/796BIJb9AlES

Walt Disney Home Entertainment. “Flamingos” In *Fantasia 2000*. Eric Goldberg, director. Burbank: Buena Vista Home Entertainment 2000. DVD, 74 min.

Warner Bros Pictures. *Bugs and Daffy's Carnival of the Animals*. In *Looney Tunes Golden Collection. Volume Five*. Chuck Jones, director. Burbank: Warner Home Video, 2007. DVD, 417 mins.

かんぱり.【プロジェクトセカイ】 限りなく灰色へ EXPERT All Perfect. YouTube Video, 2:20. https://www.youtube.com/watch?v=Px-LtgKICnY.

プロジェクトセカイ カラフルステージ! feat. 初音ミク. シネマ / Vivid BAD SQUAD × KAITO. YouTube Video, 1:19. https://youtu.be/8m1cUgLepmY?si=GmO1evPFbLgF4A18.