A Project entitled

Fusing diegetic sound into adaptive music: A compositional approach

submitted by

To Kiu Fung

A Project submitted to The Education University of Hong Kong
for the degree of Bachelor of Arts (Honours) in Creative Arts and Culture and Bachelor of Education (Honours) (Music) (Co terminal Double Degree)
in **April 2022**

Declaration

I, *To Kiu Fung* , declare that this research report represents my own work under the supervision of the *Assistant Professor, Leung Chi Hin*, and that it has not been submitted previously for examination to any tertiary institution.

Signed ___ ___

To Kiu Fung

Date: 10/4/2022

Abstract:

Music has been one of the most important factors for gamer immersion when it comes to music in gaming. Due to the nature of the player's action being unpredictable, the adaptive music is created to react and change according to players' actions dynamically. Despite its first appearance in the 80s, it is so effective that many game music composers in recent decades still make use of the system in their compositions. There are also studies exploring adaptive music in terms of composition techniques, technology and effectiveness, but most of them seem to have the premise that adaptive music is either diegetic or non-diegetic and with a rather clear boundary. This leads to the notion of the paper which is an attempt to take diegetic sound as a part of adaptive music. Although the current adaptive music system would be implemented into almost all sorts of games smoothly, there is still room for development. In the past, there have been multiple successful attempts to incorporate ambience or environmental sounds by action RPG (action role-playing game) composers. However, these diegetic elements that are fused into their compositions are mostly treated in the context of instruments exclusively and do not establish an authentic connection with players, objects or events. dynamically. This study aims to explore the possibilities of how diegetic sounds could be fused into adaptive music composition in a dynamic base that could be considered adaptive. In this study, a method will be established regarding this notion and an experimental composition will be composed with the usage of diegetic sound. Taking the diversity of game genres into account, this study will be focused on action RPGs.

Content Page

1. Introduction

1.1 Contextual background

Different from linear media music such as film and pre-recorded television shows, disregard linear games, game composers most often do not have full control of the game progress. For the sake of making background music in sync with the unpredictable player's actions, providing a better gaming immersion experience, composers tend to utilize the adaptive music system in their composition to overcome the challenge. This music system is being used in some of the most successful RPGs (Role-playing games) such as *The Legend of Zelda series, Monster Hunter series, and NieR series.* In recent years, the quality and standard of game music have also been rapidly raised to the Hollywood film music level.

Diegetic sound, most often, is used to portray the situation of a fictional world. The more it resembles sound in real-life experiences, the more realistic the gaming content. In most cases, diegetic sound such as the sound effect of a character walking, the dialogue between characters and the environmental sound etc. usually will not be considered as part of the background composition during game development. This led to one of the purposes of this essay, to investigate the kinds of diegetic sounds that would possibly work with adaptive composition. The demand for middleware application sound engines will rise in order to implement this idea in video games. Yong (2012) explained the following:

> Soon, new data input methods will allow for even more adaptability. Hardware will be able to monitor both game and player states and will tailor the musical experiences to match. As we move ever closer to augmented and virtual reality experiences, dynamic scores will play an even bigger role in spurring user immersion. (p. 48)

With the advanced technology, there are more and more middleware applications and sound engines nowadays that are solely designed for this purpose to achieve what Yong mentioned.

Research has been done on how adaptive music affects players' gaming experiences. There are also studies on composition techniques and guides out there for composers to approach adaptive music. One of future research direction of these papers is to collect more data from the responsible game and interpret them for the current adaptive music system in real-time. This also prompts the main idea of this study.

1.2 Statement of Issue

Fortunately, I was born and raised in one of the most developed cities in the world - Hong Kong. On most occasions, accessing different video games around the world is not an issue. I was a gamer beforehand I was a composer. Since my childhood, I was a huge fan of video games. Some of the earliest games that I have experienced are *Little Fighter 2 (LF2), Age of Empires II: The Age of Kings, Mega Man Xtreme, The Dungeon Siege series and TalesRunner online.* Since receiving formal music education, I have been more focused on game music compositions when playing video games. This cultivates my hobby of listening and dissecting game compositions. Not only does this experience enables me to understand and analyze game music as a composer, but also from a gamer perspective.

1.3 Aim of the Research

The aim of this essay is to investigate the kinds of diegetic sounds that would possibly work with adaptive composition and explore how would these diegetic sounds would become a part of it.

Additionally, a guide for fusing diegetic sounds to adaptive composition will be developed for other game composers and for further research.

1.4 Research questions

1. What are the diegetic sounds that would possibly work with adaptive composition?
2. How to fuse these diegetic sounds in adaptive composition?

2. Literature review

2.1 Adaptive music

Adaptive music is sometimes called dynamic music or interactive music, but in this study, the term adaptive music will be adopted. Regarding the nature of modern video games are nonlinear, adaptive scores will be fairly different from conventional linear media. Composers are required to write music that would manipulate moods, alter sections, or even abruptly cut out at any time (Marks & Novak, 24). Adaptive music captures the idea of responding to players' actions in music. Typical video game states are influenced by intensity such as the number of enemies, winning and losing, safe and danger, stealth and detection, rewards, health bar, and many other variables (Young, 2012). These conditions and circumstances are all triggered by the player both directly and indirectly. In relation to players' input, music is very often dissected into various parts to achieve flexibility.

Empirical studies have been made regarding the effectiveness of adaptive music uses in-game based on its existed theory. Results show that music is of paramount significance for players' experiences. In addition, adaptive music demonstrated a better gaming experience perceived by players when compared to its linear counterparts (Plut & Pasquier, 2019). In terms of music functions within games, Sweet (2015) stated that there are seven major aspects for music to be a compelling and useful device to bring players into the game. They are, set the scene, introduce the character, signal a change in game state, increase or decrease dramatic tension, communicate an event to the player, emotionally connect a player to a game, and enhance narrative and dramatic story arcs. These functions of music are providing a clear direction of how to measure the effectiveness of adaptive music in video games, especially on how my study will enhance or

strengthen certain kinds of aspects. As a large portion of the music used in the adaptive music system is non-diegetic music, this study will also focus on how to fuse diegetic sound into diegetic music as the first step when examining the adaptive music system.

In working with adaptive music systems, there are two major approaches employed by the composers which are vertical re-orchestration and horizontal re-sequencing (Young, 2012). Although there are other approaches other than the two previous ones, it is discerned that they are not widely exploited in RPGs. Regarding the fact that vertical re-orchestration and horizontal re-sequencing are more common approaches to action RPGs in recent decades, this study will also be limited to these two approaches. Although in some cases, it is nearly impossible to distinguish which approach is being used simply by ear, these two techniques will be taken into consideration in this study convoluting the feasibility of the final product.

2.1.1 Vertical re-orchestration/remixing/layering

Vertical re-orchestration is one of the major studio-based adaptive composition techniques that involve the addition or removal of distinct musical elements or melody lines to the currently playing piece. This strategy enables seamless transition of scenarios by fading entire instrument lines in and out according to players' activity in the game, such as location, weather or combat, etc (Cutajar 2020). Vertical re-orchestration adopts a musical groundwork that is typical emotionally cryptic and then features various rhythm elements, rhythm, tempo, or instrumentation so that they would be transited to each other seamlessly in real-time (Young, 2012). In music theory, "vertical" is to indicate a simultaneous event such as the sounding of several concurrent notes in a chord. Vertical orchestration may be also coupled with the concept

of “interactive stems”. It might be confusing for some who are relatively familiar with mixing engineers. However, in this field, the word “stem” is a very different idea. In audio mixing, recordings of several isolated parts of a piece are regarded as a stem. It could either be an individual instrument or it may be a group of instruments that form a certain percentage of the whole ensemble (Phillips, 2014). In the video game development community, “layers” are interchangeable with “stem”. However, there are fundamental differences between them. In adaptive music, vertical layering involves the playing of multiple independent audio files simultaneously within the game’s audio engine. Layers are either stacking on top of each other in perfect synchronization. This is far different from the process of simply preparing stems for mixing as the nature of the interaction is absent. Interactivity is achieved through the independent manipulation of the layers, enabling the overall track to change in accordance with the fluctuating state of the game (Phillips, 2014). An excellent example would be the adaptive music in *The Legend of Zelda: Skyward Sword* when players travel from shop to shop. Adaptive music in *Skyloft* features the usage of vertical re-orchestration as it shifts the tone of the same pieces are based on players’ location (Nintendo, 2011). One of the most obvious features of this technique is the playback line of the music usually will keep going not “jumping” around like Horizontal re-sequencing. This technique is mostly used in alternating fluctuations in intensity in different game states.

The approach of vertical re-orchestration may sometimes be called vertical remixing or vertical layering as the concept of the approach is to add or take away layers of music to create the level of intensity and emotion coordinated to the gameworld (Sweet, 2015). As these terms are simply

interchangeable, the study will simply use the term vertical re-orchestration in a later passage as it seems to be a more used name among other researchers.

2.1.2 Horizontal re-sequencing

Horizontal re-sequencing is another major technique in adaptive music. Horizontal re-sequencing is also called cue-to-cue transitions as it describes the scheduled music change depending on the kind of certain events that are called (Cutajar, 2020). The compositional approach of this technique involves writing music that is capable of shifting different sections in any given order, at any time without being noticed by the player. This approach is often entangled with multiple music sections or pieces. In essence, a diverse technique of transition is being implemented along with horizontal re-sequencing to sort of "glued" them together (Young, 2012). Some of the most notable transition effects are crossfading, abrupt cut, and adding stinger. These methods of transition are crucial to horizontal re-sequencing as its idea lay in how well is the music transit to others without being noticed by the player. The origin of the device of horizontal re-sequencing could be traced down centuries ago. In Mozart's musical dice game, pieces are broken down into segments consisting of the contents of a single measure. They are assigned numbers and the order is determined by the dicing result. Mozart composed the segments so that they could be recombined in an immense number of combinations. His game is, in fact, a low-tech but mathematically complex demonstration of a horizontal re-sequencing method (Phillips, 2014). One of the most notable examples of horizontal re-sequencing utilizing abrupt transition is when the player enters the hub from the village (Monster Hunter Rise, 2021).

Horizontal resequencing allows the music to wait for the right moment of transition between pieces. This is often achieved by waiting for the completion of the measure or musical phrase of the ongoing soundtrack to finish and fade out. One downside to this technique is the way that the triggered musical change will always lag behind the visual change, even if by a small amount (Cutajar, 2020). Although it is to determine whether or not it is a true "downside" as this feature could also provide a more obvious but gradual change of mood for the players. It is arguable that this sort of delay could be making good use of some specific situation in games.

According to Young (2012), horizontal sequencing can also have different orchestrations, allowing for even short transition cues to have several different instrumentation possibilities while vertical compositions can also have different transitions triggered overtop at any point as well. A feature of this technique is the playback line of the music usually will be "jumping" around visually in sound engines or DAWs. This technique is used mostly used in changing between different pieces or short music blocks. It is important to note that it is possible for composers to achieve the same result by using either vertical re-orchestration or horizontal re-sequencing. However, it would be most effective when using them according to the properties of the composition.

2.2 Diegetic and non-diegetic sound

The term diegetic would be traced back to its origin which is from the book, *The Republic* by Plato. He introduced the term to describe any type of narrated entertainment. In a restricted sense, diegetic and non-diegetic refer to the source domain of a stated entity: diegesis or extra-diegesis (Berndt, 2011).

Diegetic sound means the audio which occurs within the game itself and therefore can be heard by the player character (Cutajar, 2020). These sounds are real within the gameworld and they signify events or information that is real in the game (Ekman, 2005). A sound that has an identified source within the diegesis is classified as diegetic sound. Usually, the source is either visible or implied in the narrative. Examples are the sound of the footstep of a character, wind blowing, sea breeze, etc. In the category of diegetic sound, there is diegetic music. It refers to music that is performed within the scene (Berndt, 2011). For example, the character entered a restaurant that has a live band performing music or the music that the character is listening to from his/her headset. Either way, the music would be heard by the character thus they are considered to be diegetic.

Non-diegetic sound refers to audio that occurs outside the game that the player's character or the NPCs around the gaming world would not hear (Cutajar, 2020). Here, the sound effect is shown to signal an event that is not real within the gameworld (Ekman, 2005). Examples are the sound of the interfaces menu, etc. Sounds that dwell outside of the narrative world are non-diegetic. Below non-diegetic sound, there is non-diegetic music whose source is visible in the narrative, or music whose source is implied to be present by the depicted action (Thomas 2017). In most cases, they will be combat music, village theme or external background music. The traditional difference between diegetic and non-diegetic music is that the former is typically performed visibly. Usually, when listening to diegetic music, we can perceive the music to be emanated from a playback device or see a group of musicians performing. This is in contrast with its

counterpart which music is neither perceived by the characters nor is performed invisibly (Phillips, 2014).

The mechanics of the diegesis (world simulation, in a sense) that create or output is diegetic. If the further output (for example, interface sounds or the musical score) is declared non- or extra-diegetic (Berndt, 2011). Diegetic and non-diegetic sound could be also called diegetic audio and extra-diegetic audio. These two sets of concepts are basically interchangeable and this study will adopt the former.

Given the fact that diegesis is a concept derived from cinema, literature and theatre, the terms diegetic and non-diegetic are found to be problematic in interactive media such as video games. This often appears when a musical cue warns of the upcoming danger in RPGs. In a strategy game, when a work unit is selected by the player, the unit might proclaim "More work?". And in *Diablo II,* an action RPG, when the inventory is full, the avatar will proclaim "I'm overburdened". In the first impression, it might be identified as diegetic due to the usage of first-person personal pronouns. However, it might also be interpreted as non-diegetic since it is unclear whether or not the avatar is talking to (itself or the player) and, at the same time, providing information about the game system (inventory in this incident) (Jørgensen, 2011). As players are most likely to adapt their playing behaviour according to this non-diegetic information, this can be influential for the diegesis. The confusion is that the sound in these examples has a double status in which usability information is given to the player and, at the same time is stylized to fit the fictional world (Jørgensen 2011).

To clarify, the narrow sense of the terminology is used in this study. That means only the source domain will be considered rather than the range of influence together.

2.3 Concrete sounds and abstract sounds

Concrete sounds and abstract sounds are two terms used to describe the sounds based on the way they are produced in games regarding sound design suggested by Mehmet Can Uzer (2016). Sounds that are analogous to real-world occurrences are referred to as concrete sounds, while sounds that are produced to describe imagined, fictional, or metaphorical occurrences are referred to as abstract sounds. Serious processing of recorded concrete noises and synthesized audio is common in designing abstract sounds. Moreover, they also lack direct correlation to sound sources in real life (Uzer, 2016). As abstract sounds are created from the imagination in response to a certain design goal, they are considered symbolic or metaphorical.

3. Methodology

In this section, the first part will be a research on the most used types of diegetic sounds used in video games and the existed cases of using diegetic sound in video games' background compositions. Taking these into consideration, a list of the most used diegetic sounds in games will first be created for further analysis and research. Then these diegetic sounds will be classified according to their nature using *Dynamic integration,* a model proposed by Mehmet Can Uzer (2016) as an approach to diegesis and sonic interface in game audio. Following the grouping results is an analysis examing which aspects are these groups of diegetic sound would be used as a musical element in an adaptive composition. This is based on the expectations of how well they would be immersed in the universal western music theory system in terms of pitch, rhythm, timbre, etc. These cases will be focused on action RPGs including the *Monster Hunter series*, *series*, *Kingdom Come: Deliverance, the Witcher series, the Assassin Creed series, The Legend of Zelda series, Final Fantasy series and Black Desert Online.*

The second part will be an original adaptive composition with an attempt of applying diegetic sounds in it. In addition, the composition will be an experiment on mixing the two existing adaptive music techniques, vertical re-orchestration and horizontal re-sequencing. This is to give greater flexibility when applying diegetic sound into own composition. A simplified guidance system is expected to be developed to assist the composer in determining what diegetic sound to apply in composition for various scenarios.

4 Results

Upon research, here I have concluded five of the most used diegetic sounds in *Monster Hunter series*, *Nier series*, *Kingdom Come: Deliverance, Witcher series, Assassin Creed series, The Legend of Zelda series, Final Fantasy series, Black Desert Online.*

1. Dialogues (between players and NPC/ NPC and NPC)
2. Diegetic music (band performance in a restaurant, music from radio)
3. Ambient sounds (atmosphere)
4. Active sounds (footsteps, weapon slash, the collision of armour pieces, magics/spells, avatar's grunt, monster grunt, doorbell, etc.)

In this study, the terms external and internal are interchangeable with diegetic and non-diegetic. To avoid confusion, the integrated group is not classified as diegetic or non-diegetic and not be considered in the composition. The underlined text in the model is where the adaptive music takes place. It is considered external because most adaptive music is non-diegetic music. In general, it would be affected indirectly by the player's action, such as the location of the player. The four coloured texts in the internal row are all considered to be the most used diegetic sounds in action RPGs in the above investigations. Static ambience (linear loops) is rarely used in open-world action RPGs as players tend to travel around and explore different areas during their game progression. Therefore, this study will focus on its counterpart, adaptive ambience. As for the narrative NPC comments, it is mostly applied to cutscenes whose nature is linear. Thus it is not included in the above diegetic list and will not be one of the research directions. Active sound is a term that I suggested specifically for this study which will be further elaborated on in

the next section. Although the dialogue is in the same slot as other active sounds, it deserves to be considered separately in this study as it would stand alone as the main element of the music in the adaptive music system. Noted that dialogue is not classified as a member of active sounds. Other than active sounds being a bigger group and dialogue being its own group, the highlighted diegetic sounds above are corresponding to the below sound groups in the below table. In the next section, I will analyse and discuss any existing case of using the above diegetic sound in the adaptive music system. Then a composition will be produced as an attempt to fuse the chosen diegetic sounds into adaptive music.

Dynamic activity→ Narrative Integration↓	Static	Adaptive	Interactive
External	- Static music (linear loops).	***Adaptive music (that is affected by the changes in the setting).***	- Interactive music (that is triggered by player input) - External narration (narrator's voice) - Overlay UI (menus)
Integrated		- Internal avatar sounds (narrated character thoughts, heartbeat)	- Integrated UI (using boosting items like potions, collecting bonus items like mushrooms and chocolate bars, purchasing items at kiosks)
Internal	- Static ambience (linear loops) - Static source music.	- Adaptive ambience (temporal changes in the environment) - Narrated NPC comments - Aired sound (radio announcements).	- Reactive environment - Dialogue - SFX (weapons, footsteps, gas pedal) - External avatar sounds (grunts) - Aired sound (changing the radio channel)

Figure 1. Dynamic integration of game audio with examples (Uzer, 2016)

5. Analysis and Discussion

5.1 Dialogues

Dialogues between players and NPC/ NPC and NPC are usually a narrative device to depict the gameworld and story plot. Since clearance is of paramount importance in language to be effective and it would be rather unnatural to consider fusing them into adaptive music. Additionally, the language they use might be fully computer-generated or made up by game developers. Such as the “Chaos language” in the *Nier series.* In that scenario, players are forced to focus on the subtitle in order to understand the content of the dialogue. Moreover, plot dialogues often appear once according to the normal flow of time, disregarding intensionally restart and save loading action, it is possible that the player will miss out on some of the information if he/she does not pay enough attention to the dialogue. Some might argue that in most action RPGs, when you revisit the NPC after talking to them, for example receiving a quest, they are very likely to repeat some of the most important information. However, this is usually a feature to prevent the player from getting lost in terms of progression and is simply an insurance for the player who fails to catch the important information in the first encounter. Thus there will not be an identical replica of the previous dialogue, and most often, it would be a brief round-up that only gives the most crucial information regarding the first encounter. As almost all the time we want the players to understand the dialogue as soon as possible, and not be using the revisit option, it should be kept neat all the time. Some games might render players the ability to perform time travel which they can experience the identical dialogue back in time, but normally, this is not meant to be performed for a better catch of the dialogue. To minimize the cognitive load (the efforts involved in processing multi-modal information and use of working memory) of

the player when receiving information, it is concluded that most of the standard dialogues are not suitable to be used as a musical element in an adaptive composition.

However, there may be exceptions in cases where the character, including the player's avatar and NPCs, has a very musical personality and background setting that is designed to talk or present itself musically. For example, a race that is designed to be communicating with each other through singing or talking in a certain kind of rhythmic pattern. In this setting, it will be suitable to fuse these dialogues into the adaptive composition. However, to this extent, it should be considered as lyrics rather than dialogue. Given the above background, simply writing a song-like or opera-like music excerpt will be the proper way to fuse these sorts of non-diegetic sounds into adaptive music. Although, in theory, it is completely achievable, game developers of some of the most successful action RPGs such as *The Legend of Zelda: Breath of the Wild* and *Monster Hunter: World* actually do not have all of their NPCs a completely recorded dialogue due to the limited time and resources. Disregard some important cut scenes that are vital to the plot or player's progression, most often when players talk to an NPC, there will only be a simple greeting word such as "Hi", "Hello" or "Greetings". Then the player will be forced to read the subtitle to understand the sentence as there is no actual recorded dialogue. As this case is purely created from my imagination, and it is not easy to find any reference or example that a race will "speak" in such circumstances, it is quite rare for it to appear in action RPSs. With the rarity of the encounter, limited time and resources in game development, and the concern for dialogue clarity, there will not be a composition dedicated to this diegetic element in this study.

5.2 Diegetic music

Due to the nature of diegetic music is much more musical than other diegetic sounds, it would be possible to directly consider it as part of non-diegetic music. Diegetic music is often synchronized with a visible sound source. That means the player would easily recognize the sound source when it is onscreen. This would be a live performer with an instrument or a radio, etc. However, the type of sound sources does not necessarily have to be analogous to real-world instruments or sound devices. Distinct from the non-diegetic music commonly used in adaptive music, diegetic music is treated as an actual sound event in the gameworld. Therefore, it will only be heard once the character is near enough to the sound source. Generally, diegetic music will only be present in a specific safe zone area where the player has the opportunity to stop and take his/her time to enjoy the music. Occasionally, the player may have direct control of when the diegetic music begins and finishes. For example, in *Ghost of Tsushima,* the player has a shakuhachi as an accessory equipped in the character's gear. The player is given the power to change the weather by playing different shakuhachi pieces. These pieces could be learnt by the player by collecting singing crickets. Throughout the game, the player can play it at any time in the open world. In some games, the players may even be able to play the instrument with its own special interface. For example, in *The Last of Us Part II,* when the player picks up the guitar, a dedicated chords interface will appear that allows the player to control their character to perform strum or broken chords. In *Sky: Children of the Light,* the player would unlock a harp and play a range of notes in its designed interface. In *Red Dead Redemption 2,* there will be a pianist performing in the tavern and the gambling riverboat. The player would interrupt the pianist by interacting with him. On some occasions, the player may also play the piano him/herself. Most of the time, there will only be diegetic music along with the ambience with no non-diegetic music

in the background. For example, in *Assassin's Creed IV: Black Flag,* when the player set sails, the sailors on board will sing a set of sea shanties from time to time with no other non-diegetic music element. That being said, there is still an example of diegetic music syncing with non-diegetic music and is one of the best examples of fusing diegetic music into the adaptive music system.

In *The Legend of Zelda: Breath of the Wild,* there is a minstrel called Kass who travels Hyrule and sings the legends of the areas he visits whilst accompanying himself on bandoneon. In one of the safe zones, Rito Village, Kass and a group of children Ritos, a recurring race with avian features, would perform along with the ongoing diegetic background theme. This event follows the physical boundaries in the real world sound event as the player would only be able to hear Kass performing when he/she is near enough to Kass. It is discerned that there is an invisible circle indicating the maximum area that Kass's music can affect and his music would only be heard by the player within the circle. In the safe zone, Rito Village, there is a default non-diegetic background theme that the player will constantly hear whenever he/she is in the village. Normally, the main melody is mainly occupied by a clarinet and a mandolin. However, if the player is near Kass, the melody will then crossfade to the bandoneon and the voice of the group of the children Ritos. Moreover, the orchestration is clearly different as Kass and the group will completely take over the clarinet and mandolin. Whenever Kass is present, there the part of the clarinet and mandolin will be absent. Here the diegetic music (Kass and Rito group) is fused to the non-diegetic theme. Then this music will change according to the location of the player in the adaptive music system. While it is discerned that diegetic music is one of the easiest diegetic

elements to fuse in the adaptive music system, it does require a specific music source in the gameworld.

5.3 Ambient sounds (atmospheric)

Ambient sounds refer to surrounding or background noise according to the environment. The major purpose of these sounds is to add a sense of presence and a specific mood to the game. Liljedahl (2011) stated that there could be multiple purposes served by an ambient or background sound at the same time, including creating a sense of physical presence, setting the basic mood and communicating emotion and arousals. They are present not to influence player actions by giving the player specific information about objects, events or situations, and they are often not connected to specific sources in the game (Jørgensen, 2006). Most ambient sounds are instead connected to virtual sources as it does not necessarily provide an entirely faithful representation of the image in every case. That means the source of the ambient environmental sound may not be present on-screen or synced with the image. Ambient sounds may thus coney the feeling of a realistic space by presenting virtual offscreen sources (Jørgensen, 2006).

When it comes to fusing ambient sounds in music, it will usually be connected to an existed genre, ambient music. It emphasises tone and ambience rather than melody, musical structure, or rhythm in traditional music. Although with the name ambient music, the presence of ambient sounds is not necessary for this type of composition. Ambient music is prone to be a music style that emphasises tone and ambience rather than indicating specific kinds of elements therefore we may find modern ambient music also include the usage of synthesisers. It is important to note that most ambient sub-genre compositions, even those that involve ambient sounds, do not treat

them as a vital part of the composition. For example, the ambient sounds may not be constantly present, would easily be covered by other musical elements and would not have great influence if they are absent. Nevertheless, it serves as a great reference and foundation for what type of texture and timbre would be blended with ambient sounds.

As it is acknowledged that one of the major purposes of ambient sounds is to convey the gameworld's intrinsic reality and enhance players' immersion, ambient sounds should also be the axis of the composition when fusing into music, unlike other ambient music sub-genre. This type of composition will still be classified as ambient music, but instead, ambient sounds will become a vital part of the composition. In this study, this type of composition will be called ambient oriented music for convenience. The principle is to blend the ambient sounds and musical elements in terms of texture. Moreover, the ambient sounds should be as audible as possible. This would be easily achieved by using lesser instruments in which gradual sustained long notes or simple motif loops will be the major part of the musical elements. It is believed that a suitable amount of percussion would be very effective when trying to add some energy to this type of composition.

Due to the nature of ambient sounds in video games being constantly present and restricted to certain environments, the most effective way of using ambient oriented music with adaptive music would be to alter the instruments or sounds layers on top of the continuous ambient sounds. This alteration is to achieve different game intensities corresponding to different game states. In most open-world action RPGs, the two game states, *neutral* and *combat*, will be the most important part of the adaptive music system as players will spend most of their time

listening to the sound or music from these two states. A simple example would be, when the player is exploring the world and does not encounter any enemy, he/she will be listening to the music composed specifically for the neutral game state. If the player does encounter any enemy, the music will change to combat state music. When applying to adaptive music, the ambient oriented music mentioned above will be the music for the *neutral* state. Extra layers of instruments will be added to the *neutral* state music becoming the *combat* state. As a result, the music will then be altered to adapt to the corresponding game state.

There are two existing cases of fusing ambient sound into adaptive music composition. The first one is *No Man's Sky* and the second one is *Elden Ring*. In *No Man's Sky* when players fly through space or inside the atmosphere of a planet, ambient oriented music is being used in the adaptive music system. In this example, the game states will be *planet* and *space*. When the player is flying on a planet, the ambient oriented music in planet state, which is lighter and calmer, will be adopted. If the player starts to speed up their spacecraft or enter space, the music in space state will takeover. Generally, the music in *space* state is an altered version of the one in *planet* state, but instead, it will have an extra layer of synthesizer rhythmic pattern and sometimes with distortion effect. According to the different environments of the planet, the *planet* state music will also be slightly different.

In *Elden Ring*, players will spend most of their time exploring the gameworld and combat. There are basically three game states, *exploration, normal combat* and *boss fight*. The music in the first two game states is one of the best examples of ambient oriented music. Generally, the ambient oriented music in the *exploration* state consists of very few instruments. Most of the musical

elements are sustained gradual long notes in traditional or processed instruments. The instrumentation and ambient sounds will also be changed according to the region of the player. The music of *normal combat* state is the hyper version of the music in *exploration state.* Most notably, percussions and pre-existing instruments with accented rhythmic patterns are being added to the *exploration* state. While the ambient sound is a crucial element in both the *exploration* and *normal combat* states, it is decorative in the music of *boss fight* state. The music in *boss fight* state will be adapted to the boss theme (a completely different piece of music) according to which boss does the player encounter. That also contributed to the reason why the ambient sounds will be hardly notable in *boss fight* state. As mentioned before, the source of ambient sounds may not be present all the time. It is worth mentioning that when the player enters an underground cemetery, the player will hear a layer of whispering sound periodically in the ambient oriented music. This whispering sound does not have a specific source on-screen, but it will keep appearing throughout the underground progression. As there is a kind of underground enemy that makes a similar sound, it mostly implies the whispering source. This element provides a sense of narrative and uniqueness to the ambient sounds unlike the traditional one. In addition, ambience can also include abstract elements, mostly in a form of a combination of artificial sounds and musical components (Uzer, 2016). That means the ambience in the ambient oriented music would be pure imaginative or very different from the real-life ambience according to the gameworld. This provides a limitless possibility to game composers in terms of creating ambience elements. The usage of ambient oriented music in *No Man's Sky* and *Elden Ring* are two great examples of how to fuse ambient sounds into the adaptive system, therefore this study will not include a composition regarding ambient sounds.

5.4 Active sounds

Active sounds is a term that I proposed in this study to describe any organic, mechanic, object, non-dialogue and non-ambience sounds with a specific source within the gameworld. Note that diegetic music is not categorized as active sound. This sound category also includes sounds that are emanated or triggered by the PC (player character). For example, amour piece collision, footsteps and objects break by player's character. However, these sorts of sounds are too unpredictable to be orchestrated into theme music naturally therefore they will not be considered in this section. Action sounds usually inform the presence and relative location of objects and events. Such as the air sound of the funnel may indicate an armour workshop. This means they enable the player to know what is happening in locations where they are not in the immediate vicinity. While the nature of ambience sounds may always be continuous while action sounds (non-PC triggered) are usually periodic. Hence, they are predictable and much easier to control.

In action RPGs, players will encounter different safe zones for them to upgrade their gear, receive quests, save progression and prepare for future battles. Second to exploration and combat outside safe zones, players will often spend a certain amount of their game time in these safe zones. These safe zones could be outposts, villages, towns, bases, cities strongholds, etc. Larger scale safe zones or those of significant importance in the game progression are very likely to be accompanied by a dedicated non-diegetic theme correspondingly. These themes are mostly non-diegetic music and are extremely important in conveying the mood of the gameworld. Usually, these themes are the music that players would remember the most. Moreover, the gameplay intensity and alertness in these safe zones are much lower than in exploration and combat. As a result, players would perceive the environment in a much broader sense other than

just focusing on combat. These are the two major reasons why the composition in this section is centred around this kind of theme music regarding the adaptive music system.

As mentioned before, non-PC action sounds are usually periodic. In most action RPGs, they will have a fixed triggering cycle controlled by the game engine. By manipulating the triggering cycle, these sounds would be musically synchronising the background theme. In this context, rhythm and melody are considered to be two of the most suitable aspects for active sounds to fuse into the theme. Depending on the element in the active sounds in the gameworld, harmony may also be taken into account. Active sounds that have an indefinite pitch are more likely to be used as a rhythmic element, while the definite pitch counterparts may be used as a melodic element. For example, a steam pump may have its air sound accompanying the theme rhythmically like percussion, and the sound of a doorbell may be used to highlight the melody when triggered. The possibility here is immense as the active sounds from gameworlds with different settings may have their own unique abstract sounds other than concrete sounds. This provides more freedom to game composers when fusing active sounds into adaptive music.

Normally, there will be multiple areas with corresponding functions for players to enter in a safe zone. They would be weapon shops, armoury shops or taverns, etc. Regarding the adaptive music system, the composition attempt in this section will be a safe zone theme that would be altered based on which area the player enters. The active sounds that fused into the theme will also be different according to the area. This lowers the possibility of players experiencing listener fatigue, but at the same time, maintains unification of the safe zone theme.

6. Composition attempt

6.1 Game reference: Monster Hunter: World (Astera)

The composition attempt will be based on a safe zone village in an existed open-world RPG Monster Hunter: World, called Astera. This place is working as the Research Commission's base of operations where hunters, researchers, and technicians have all gathered. Most of the quests assigned as part of the story will come from various characters in Astera and travel to different facilities in the village. It is considered the most important safe zone throughout the game progression.

6.2 Diegetic elements:

Regarding the above discussion and analysis, diegetic music and active sounds will be used as the diegetic elements in the composition while active sounds will be the focus. Continuous diegetic sounds such as metal chain mechanics will be triggered frequently while their discontinue counterparts such as Cat's meow will be triggered according to the event cycle (labelled as Taec. in diegetic reference).

6.3 Adaptive music:

A new orchestral work, default theme, will be composed as the adaptive safe zone theme in the main area of Astera. The orchestration of the theme will be changed to adapt to the player's location and sometimes actions.

6.4 Product

The product will be a full-length safe zone theme composed for the main area and 3 other different orchestration versions (not in full-length) for the chosen facilities in Astera. These products will be accompanied by 4 video demos simulating the sound and music that the player will hear when travelling to various destinations. Different active sounds and diegetic sounds will be used according to the environment in different areas. The full score of the safe zone theme will be provided in *Appendix I - IV. Appendix I* is the full orchestral version while the rest are shortened versions. *Appendix I, III and IV* include a part called *Diegetic reference* at the end to indicate how the diegetic sounds are being used.

6.5 My approach to the safe zone theme

When composing for games, it is always necessary to, first, have a basic understanding of the gameworld, its story plot and the area that is accompanied by the music. For example, weather, construction styles, the history of the area and plot timeline. They often provide information about the mood that a certain area is trying to convey. It would be “grand”, “sad”, “mysterious’, etc. One of the considerations is to identify the game state of the area in which music is needed. As the game state to a certain extent indicates the intensity level of the gameplay, the intensity of the music should also be in accordance with it. The safe zone theme should always be loopable or continuous as the player tends to stay quite some time in the safe zones.

Astera is the first and most important base that the player will encounter in Monster Hunter: World. It is constructed primarily out of wood and ships once sailed by the avant-garde who reclaim the New World. It has a spectacular coastal landscape and sunny weather during the

daytime. The player should be very excited about exploring the base and preparing for the coming adventure. In addition, the music in Astera is expected to be the has the symbol of "home" that grants the player a sense of fantasy and nostalgia in the game. Besides, I also want the composition to begin and symbolise the courage of the First Man who arrived New World and built Astera. This is why the composition starts with a light solo clarinet melody. The musical idea is then being reinforced and developed in different sections, representing those who arrived at Astera following the path of the First Man. The composition is my imagination of the development progress of Astera from the very beginning to its current state where the player arrives. In this composition attempt, vertical re-orchestration is being used as its major adaptive music technique for a smooth transition when the player enters different facilities in Astera. The diegetic elements used in the composition are fully based on the environment around the player. Most of them have their own triggering cycle synchronized with the music. For example, the cat's meow is aligned with part of the piece but with its unique triggering cycle. That renders me the ability to control which bars and beats will there be the sound of a cat's meow. Other continuous diegetic elements such as metal chains and waterwheels may have a constant rhythm and fixed triggering cycle accompanying the composition. These diegetic elements work with the premise of an actual sound event with a specific sound source in the gameworld that the player will only hear once they are in his/her vicinity. Ideally, they should be also synchronized with the visual image.

6.6 Location reference

Location 1: Main area (Not inside any facilities)

Orchestration 1: Default theme (Full orchestra)

Length: Full-length

Usage of diegetic sounds: Active sounds - elevator mechanic, metal chain sound, Cat's meow, Air blow by the cat, waterwheel

Demo link: https://youtu.be/WXon7Cgzkg8

Score: See Appendix I

Location 2: Player's hideout

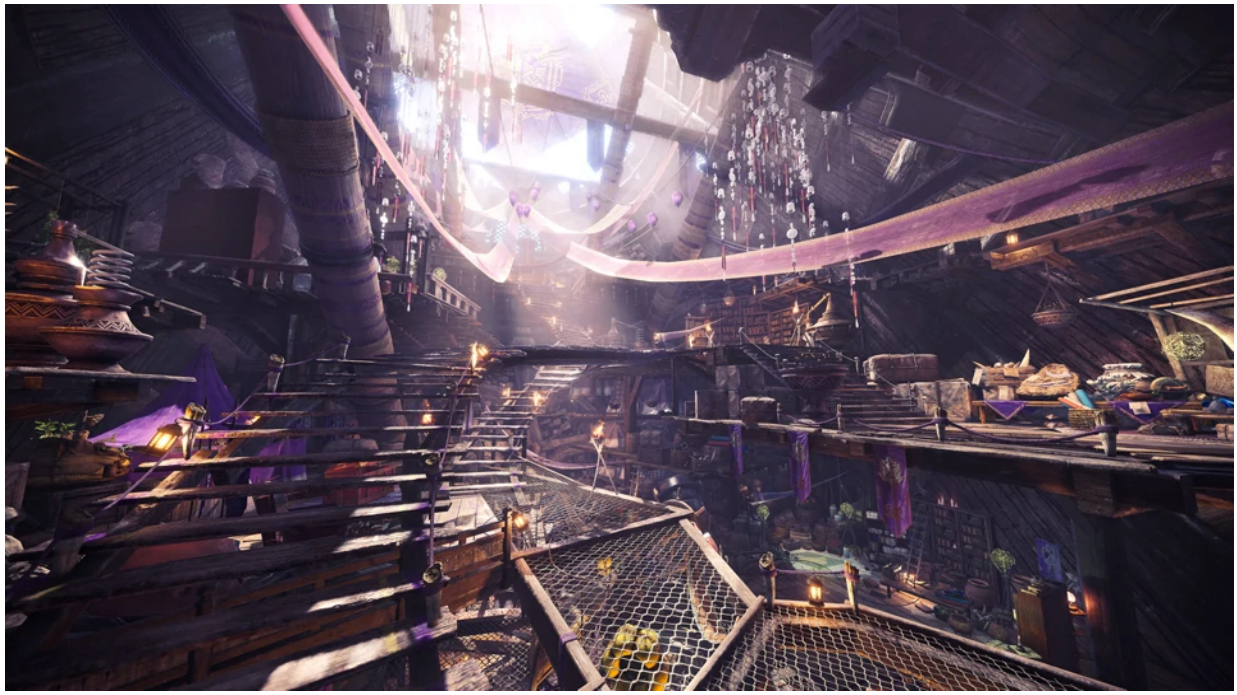
Orchestration 2: Default theme (without brass section and the harp is accounted for the main melody)

Length: Shortened

Usage of diegetic sounds: Diegetic music - harp played by the two cats

Demo link: https://youtu.be/yPIMs8NoHSA

Score: See Appendix II



Location 3: Research Base

Orchestration 3: Default theme (Pitched percussion and diminished woodwind section)

Length: Shortened

Usage of diegetic sounds: active sounds - cat's meow, hanging bell chimes, steam pump from large boiling alchemist beaker

Demo link: https://youtu.be/gzstzvsIT6s

Score: See Appendix III

Location 4: Astera's Workshop

Orchestration 3: Default theme (Brass, percussion and diminished woodwind section)

Length: Shortened

Usage of diegetic sounds: active sounds - metal mechanic, hammer, steam pump, melting

Demo link: https://youtu.be/jju8qbTIzKo

Score: See Appendix IV

7. Conclusion

From the compositional attempt, a guide called *DSAM,* standing for *diegetic sound to adaptive music*, is proposed to simplify the step of fusing diegetic sound to adaptive music in various scenarios and for future research.

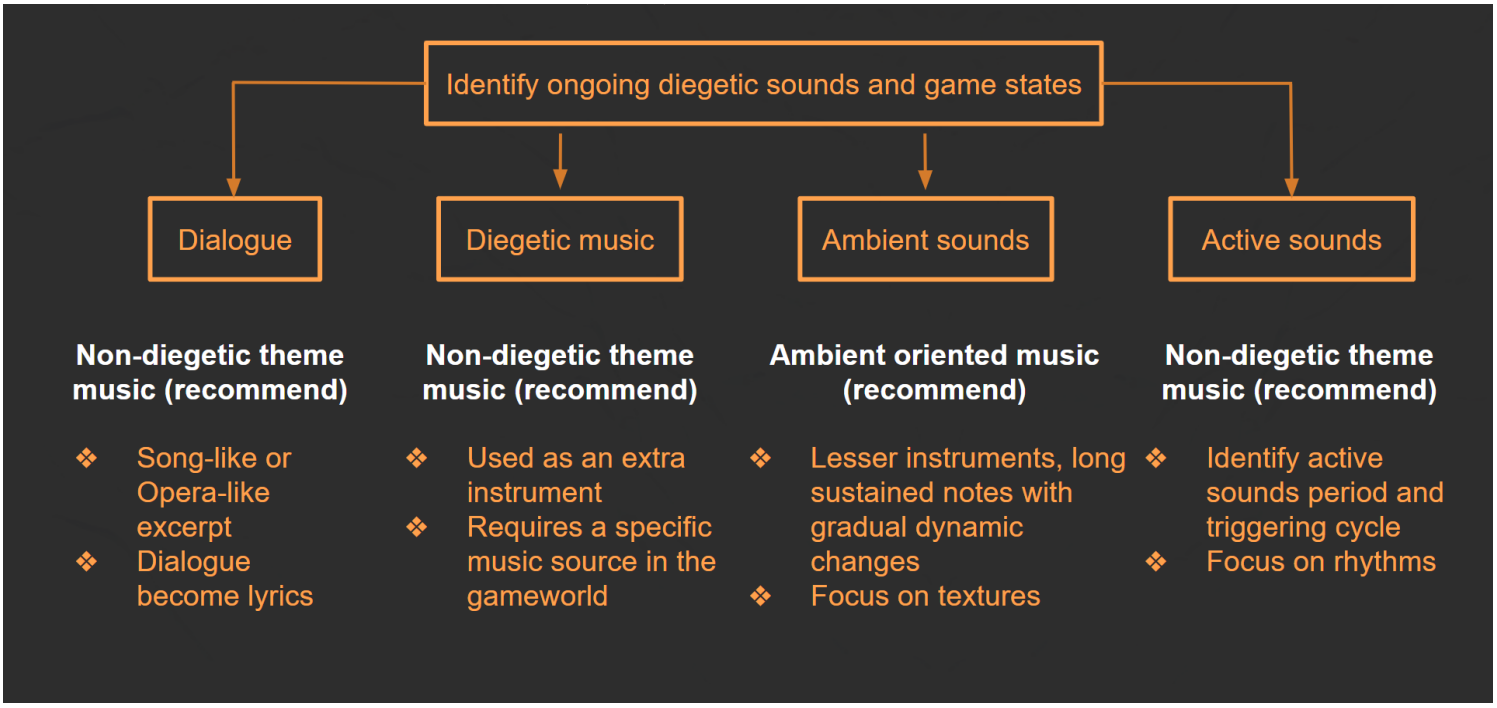


Figure 2. DSAM (diegetic sound to adaptive music)

The first step when approaching will be to identify ongoing diegetic sounds and game states. It is important to know whether the music is scored for multiple game states or a single one. Identifying ongoing diegetic sounds enables composers to classify them as the diegetic sounds mentioned above. The white texts down below different types of diegetic sounds are the recommended background music to be implemented in the adaptive system. The lowest orange points indicate how to approach these diegetic sounds musically and which aspects shall be focused on.

This paper examined the possibilities of fusing various diegetic sounds into the adaptive composition in open-world RPGs. The final product merely served as an example to support this notion. In this study, the composition attempt convoluted to the pre-scored background music as it is acknowledged to be the majority of game music in the current industry. That means other types of music that are possible to be used in the adaptive music system such as algorithmic music is not included in this research. Moreover, there are still various kinds of diegetic sounds that are considered in this study due to their properties, but this does not imply those diegetic sounds are not usable. It is believed that composers will need to have more involvement in sound design during game development in order to make good use of this approach as the diegetic sounds will also be part of the composition. There is no fixed formula of how composers would achieve the best out of applying diegetic elements to adaptive music as diegetic sounds would be formed by both concrete and abstract sounds. According to the game context, the diegetic sounds would be very different and flexible. DSAM is a guide meant for open-world action RPGs and it might not be applicable to all games, but it would act reference for composers who are trying to do something similar with other game genres. There is still are a lot to explore and the idea of fusing various diegetic sounds into adaptive composition should not be only limited to the approach in this study.

In the composition approach in this paper, multiple diegetic sounds have been applied to the adaptive composition. This causes the boundaries of diegetic and non-diegetic to become rather ambiguous when this kind of product is described as a whole. Although it seems to be possibly classified the composition attempt as “Interageted” in Uzer *Dynamic integration of game audio* model, the tag is not meant to describe it. New terminology might need to be invented to identify

this kind of music. Future research may be fusing the other diegetic sounds that did not include in this study into various kinds of music, not limited to may also be there may be a computer-generated adaptive music engine that also includes diegetic sounds during the generating progress. In that, the game and sound engine would detect and adapt diegetic sounds automatically according to the player. As a result, the music will be tailor-made for players and provide them with a unique experience. The possibility and room for development of DSAM are immense, it should not only be limited to the current state in this paper. Its effectiveness for players' immersion is yet to be determined by future research.

References

Berndt, A. (2011). Diegetic Music: New Interactive Experiences. In *Game sound technology and player interaction: concepts and developments* (pp. 60-77). IGI Global.

Brown, R. S. (1994). *Overtones and undertones: Reading film music*. Univ of California Press.

Brown, E., & Cairns, P. (2004, April). A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1297-1300).

Caillois, R. (2001). *Man, play, and games*. University of Illinois Press.

Capcom. (2018). *Monster Hunter: World.* [PC, various consoles], Capcom, Japan.

Cavia. (2010). *NieR: Gestalt/Replicant.* [Various consoles], Square Enix, Japan.

Cutajar, S. (2020). *Automatic Generation of Dynamic Musical Transitions in Computer Games* (Doctoral dissertation, The Open University).

Cohen, A. J. (2000). Film music: Perspectives from cognitive psychology. *Music and cinema*, 360-377.

Ekman, I. (2005). Meaningful noise: Understanding sound effects in computer games. *Proc. Digital Arts and Cultures*, *17*.

Hello Games. (2016). *No Man's Sky.* [PC, various consoles], Hello Games, United Kingdom

Jørgensen, K. (2006). On the functional aspects of computer game audio.

Jørgensen, K. (2011). Time for new terminology?: Diegetic and non-diegetic sounds in computer games revisited. In *Game sound technology and player interaction: Concepts and developments* (pp. 78-97). IGI Global.

Liljedahl, M. (2011). Sound for fantasy and freedom. In *Game sound technology and player interaction: Concepts and developments* (pp. 22-43). IGI Global.

Marks, A., & Novak, J. (2008). *Game development essentials: game audio development*. Delmar Learning.

Munday, R. (2007). Music in video games. *Music, sound and multimedia: From the live to the virtual*, 51-67.

Nintendo EPD. (2017). *The Legend of Zelda: Breath of the Wild.* [Switch], Nintendo, Japan

Phillips, W. (2014). *A composer's guide to game music*. MIT Press.

PlatniumGames. (2017). *NieR: Automata.* [PC, various consoles], Square Enix, Japan

Plut, C., & Pasquier, P. (2019, August). Music Matters: An empirical study on the effects of adaptive music on experienced and perceived player affect. In *2019 IEEE Conference on Games (CoG)* (pp. 1-8). IEEE.

Qin, H., Patrick Rau, P. L., & Salvendy, G. (2009). Measuring player immersion in the computer game narrative. *Intl. Journal of Human–Computer Interaction*, *25*(2), 107-133.

Uzer, M. (2016). Conceptual Design Methods for Game Sound.

Ubisoft Montreal. (2013). *Assassin's Creed IV: Black Flag.* [PC, various consoles], Ubisoft, France

Sweet, M. (2015). *Writing interactive music for video games: a composer's guide*. Pearson Education.

Thomas, C. (2017). *Composing Music for Games*. Routledge.

Young, D. M. (2012). *Adaptive game music: the evolution and future of dynamic music systems in video games* (Doctoral dissertation, Ohio University).

Zehnder, S. M., & Lipscomb, S. D. (2006). The Role of Music in Video Games.

Appendix I

Astera's Parhelion

for orchestra

To Kiu Fung 2022

A Project entitled:
Fusing diegetic sound into adaptive music: A compositional approach

A project submitted to The Education University of Hong Kong
for the degree of Bachelor of Arts (Honours) in Creative Arts and Culture

in April 2022

Flute
Oboe
Cor Anglais
Clarinet in B♭
Bassoon
Horn in F
Horn in F
Trumpet in B♭
Trumpet in B♭
Trombone
Tuba
Timpani
Suspended Cymbals
Cymbals
Sleigh Bells
Mark tree
Chimes
Glockenspiel
Harp
Mandolin
Violin I
Violin II
Viola
Violoncello
Double Bass

♩=73

mp p pp pizz

11

Fl. Ob. C. A. Cl. Bsn. Hn. Hn. Tpt. Tpt. Tbn. Tba. Timp. Susp. Cym. Cym. S.Bells M.tree Chim. Glock. Hp. Mand. Vln. I Vln. II Vla. Vc. Db.

21

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

30

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

39

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

45

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

51

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

59

rit. ♩=73

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

68

Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

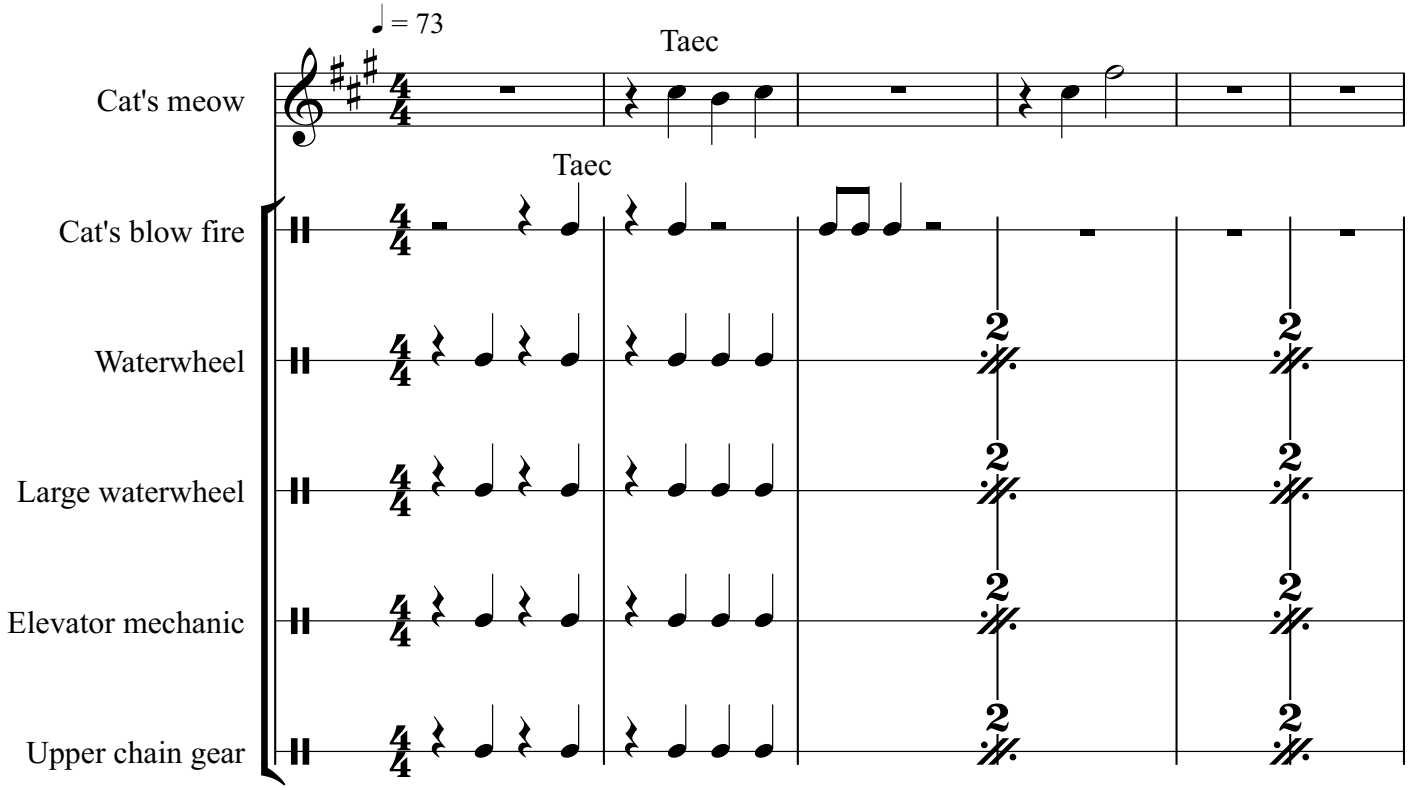
72

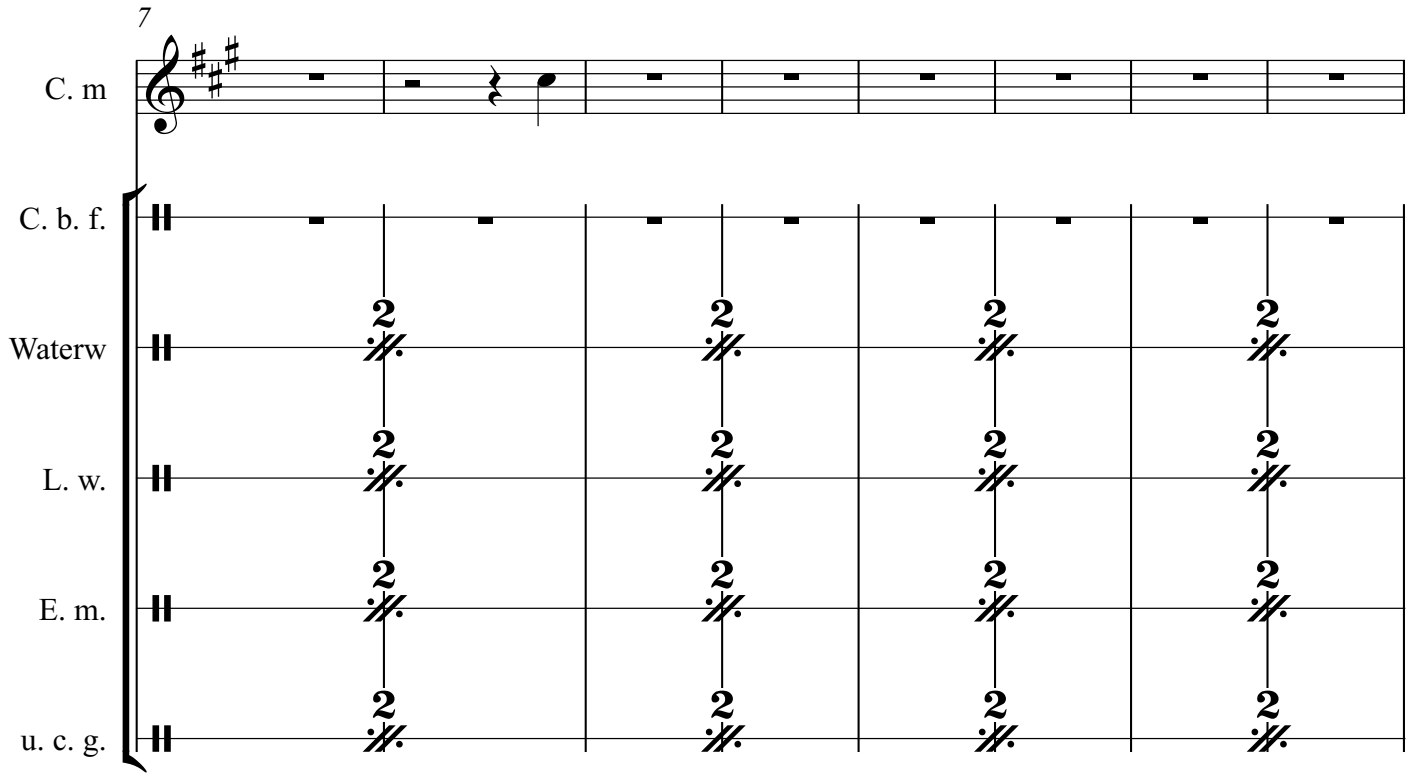
Fl.
Ob.
C. A.
Cl.
Bsn.
Hn.
Hn.
Tpt.
Tpt.
Tbn.
Tba.
Timp.
Susp. Cym.
Cym.
S.Bells
M.tree
Chim.
Glock.
Hp.
Mand.
Vln. I
Vln. II
Vla.
Vc.
Db.

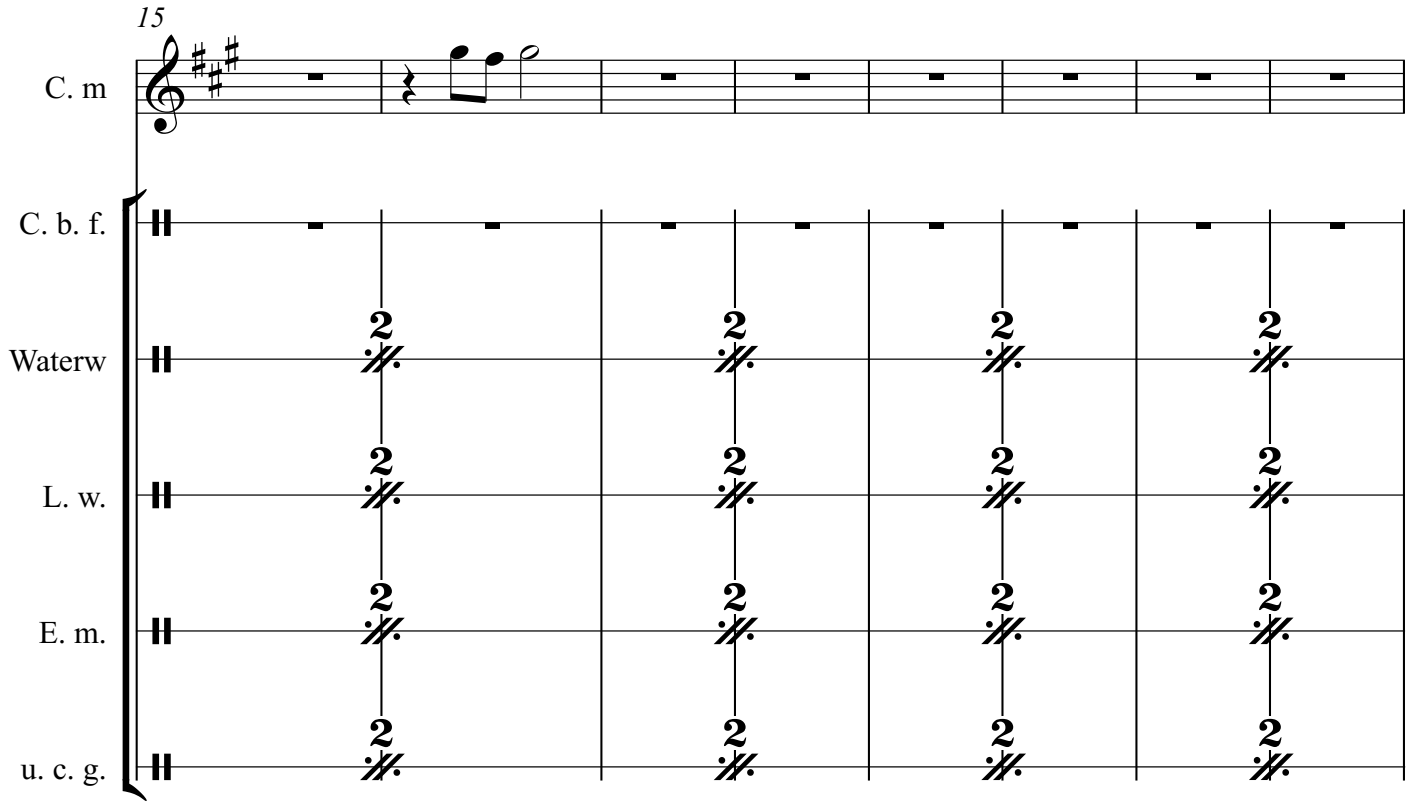
rit.

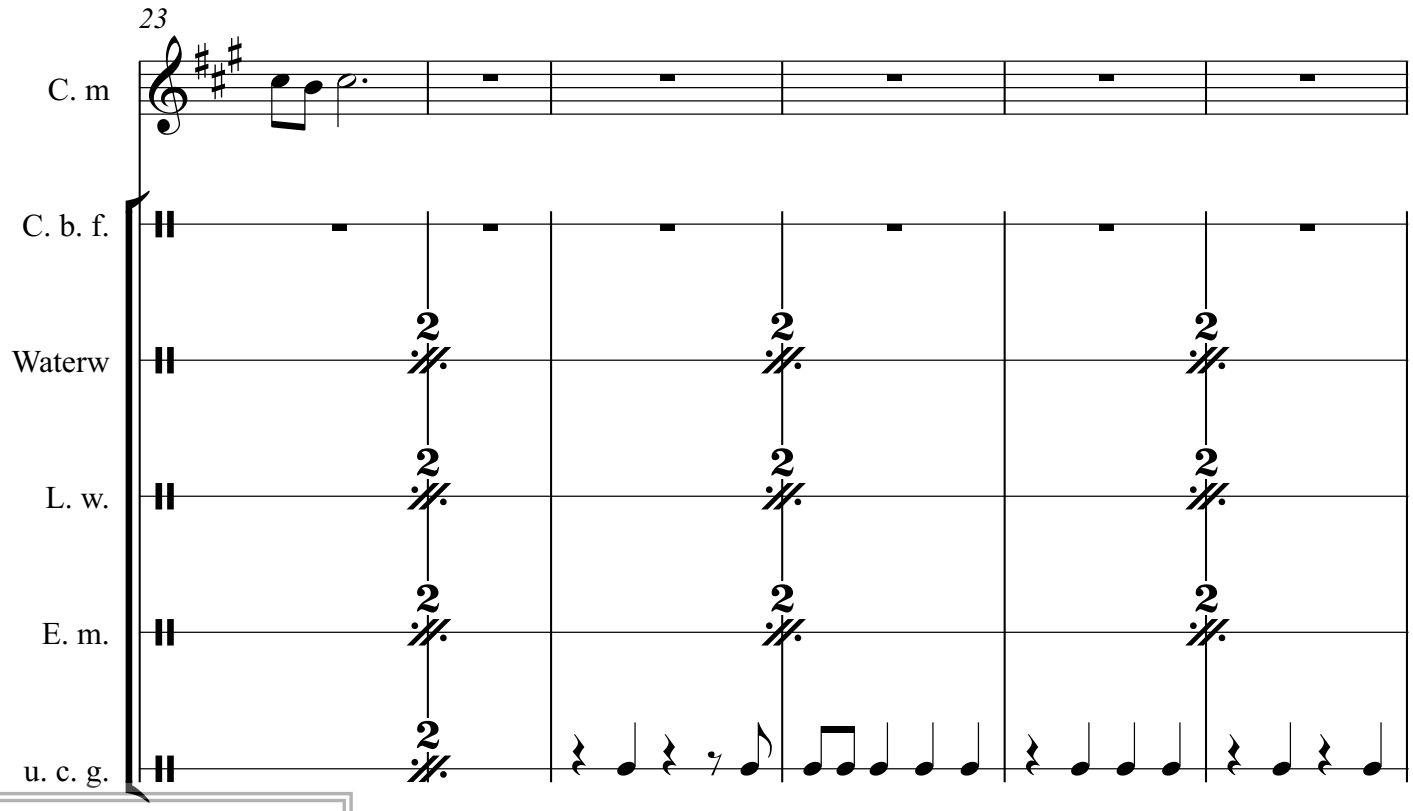
Diegetic reference - Main Area

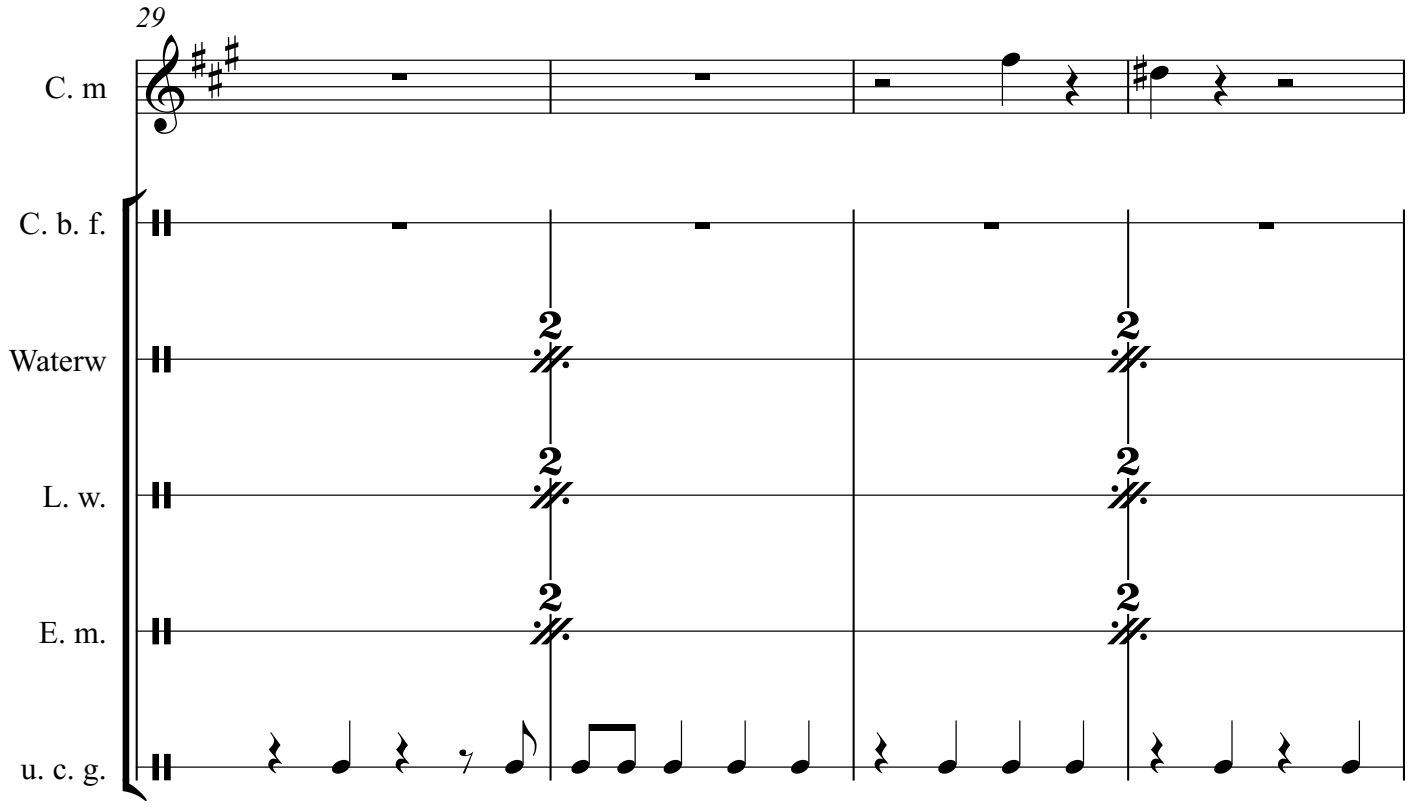
Triggers accodring to event cycle = Taec.

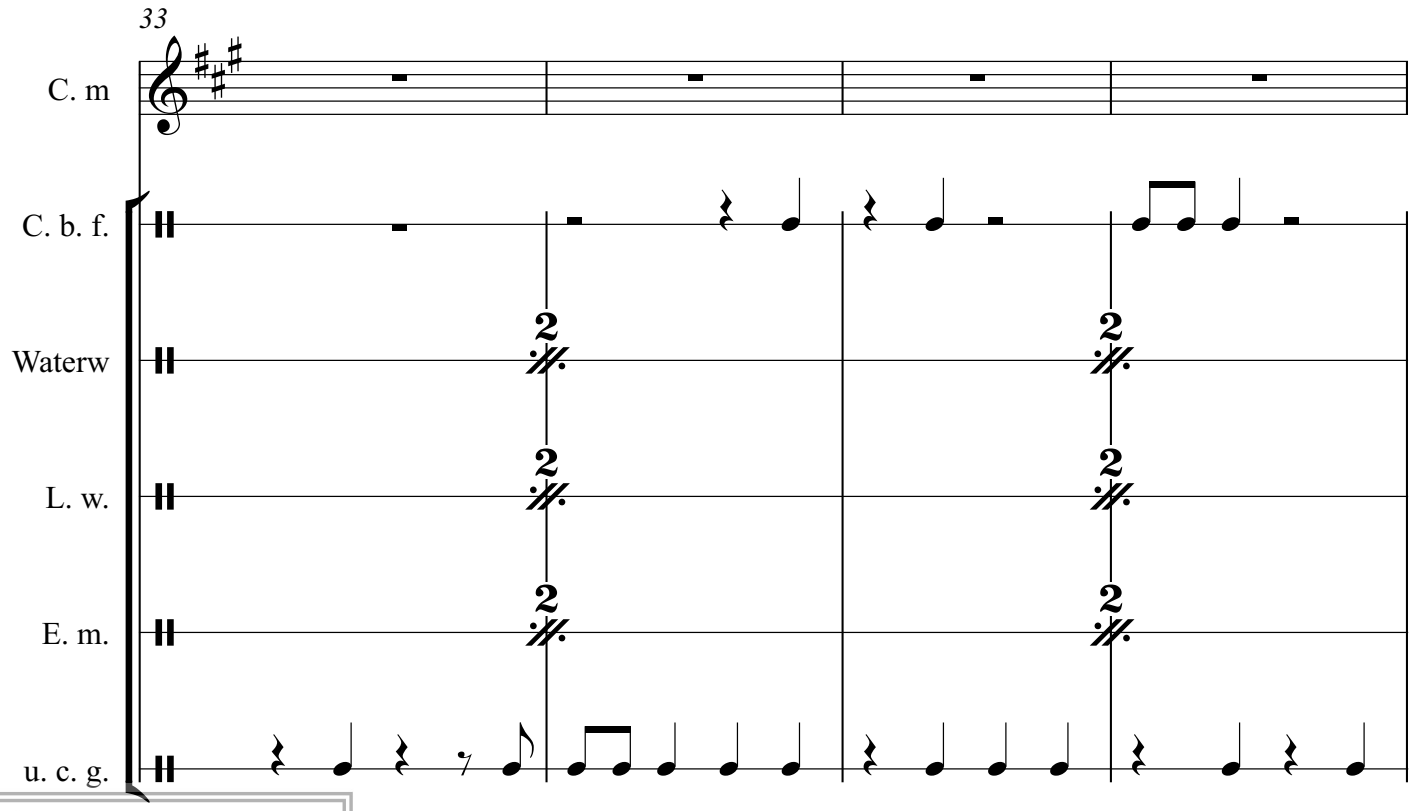


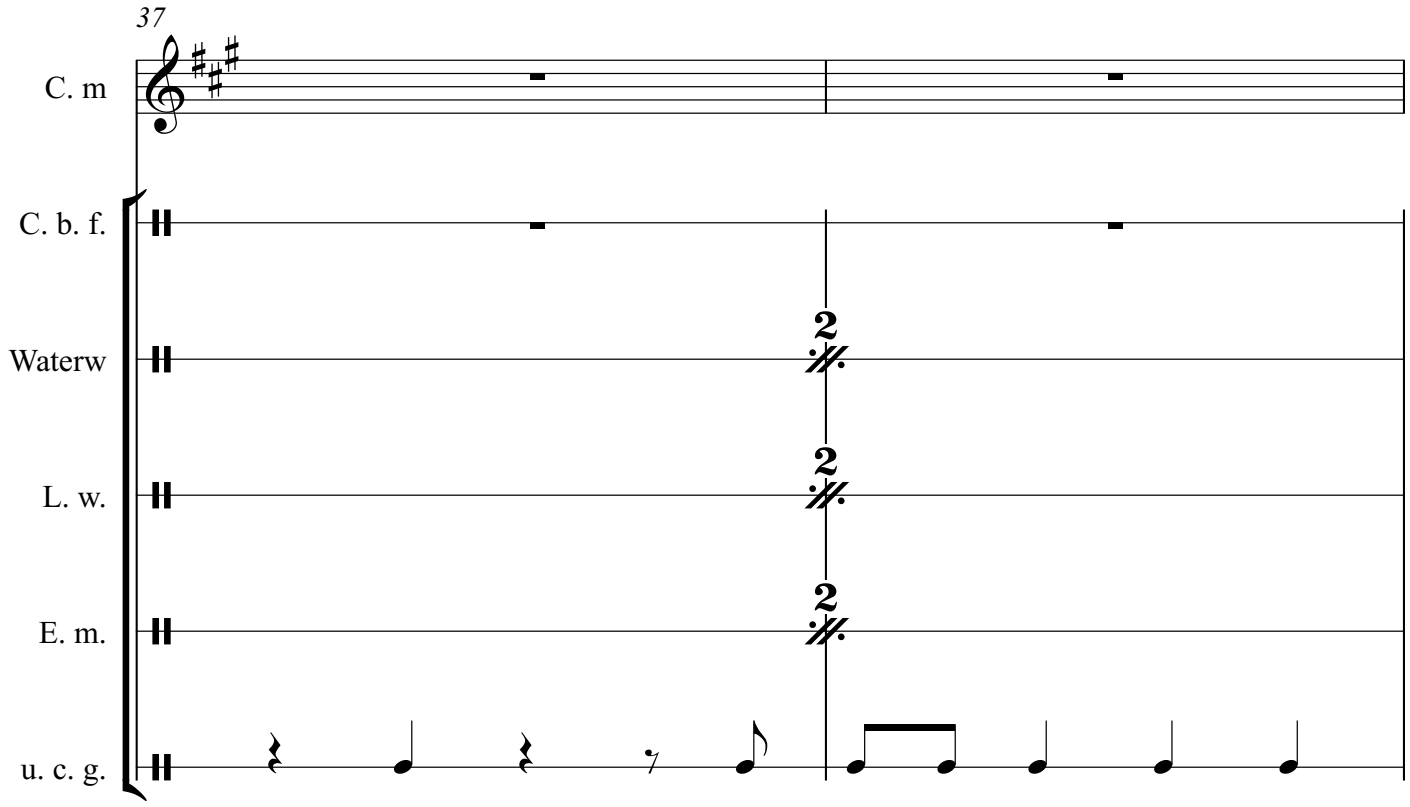


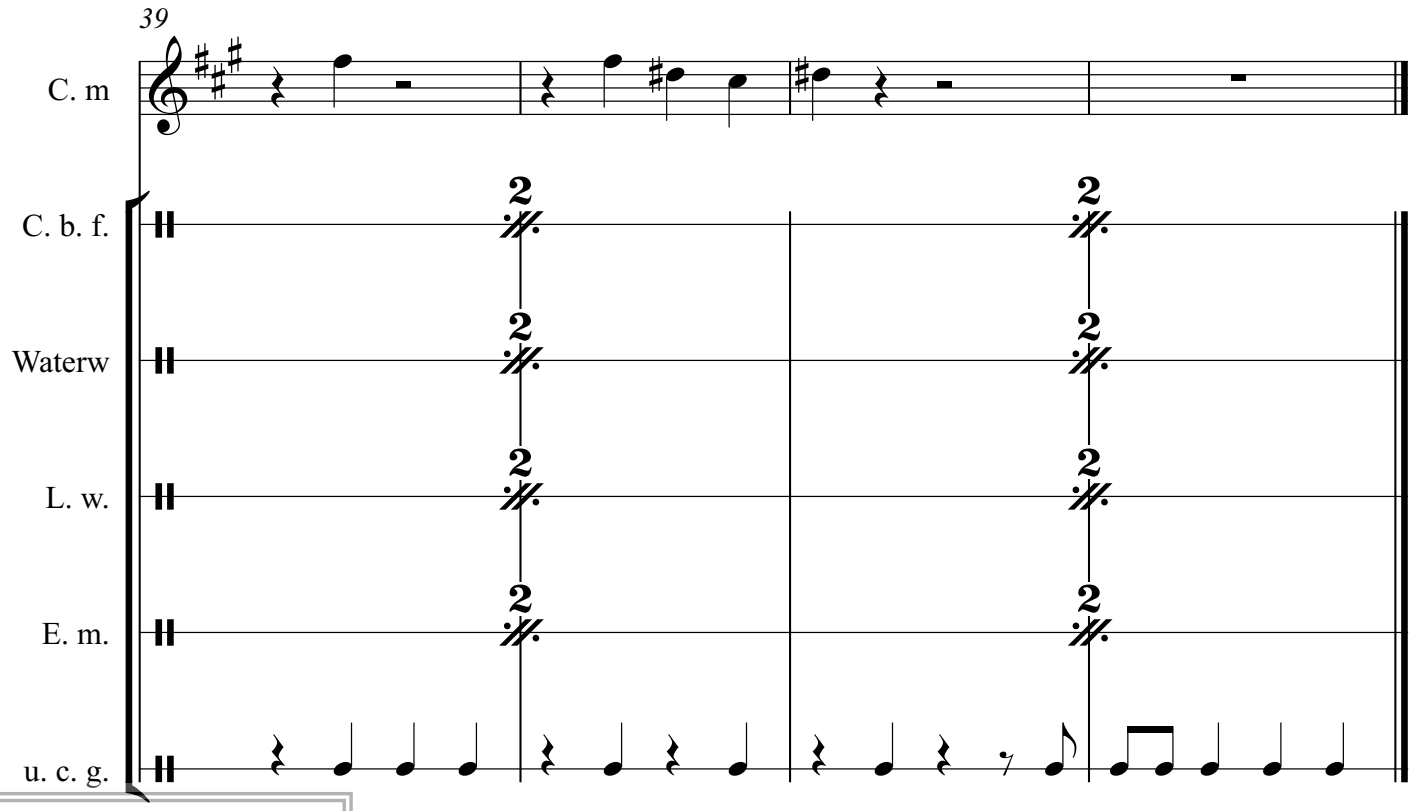






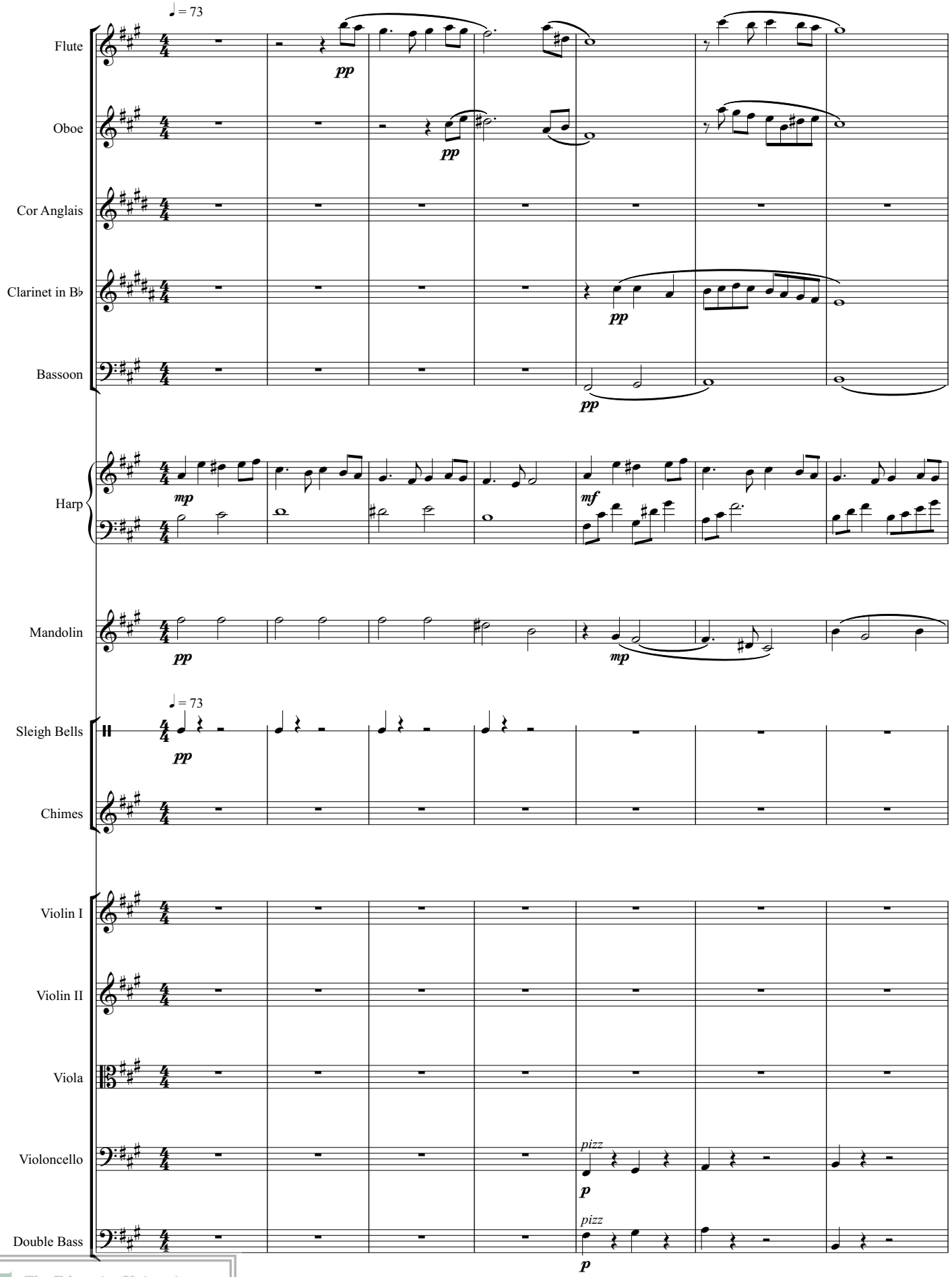


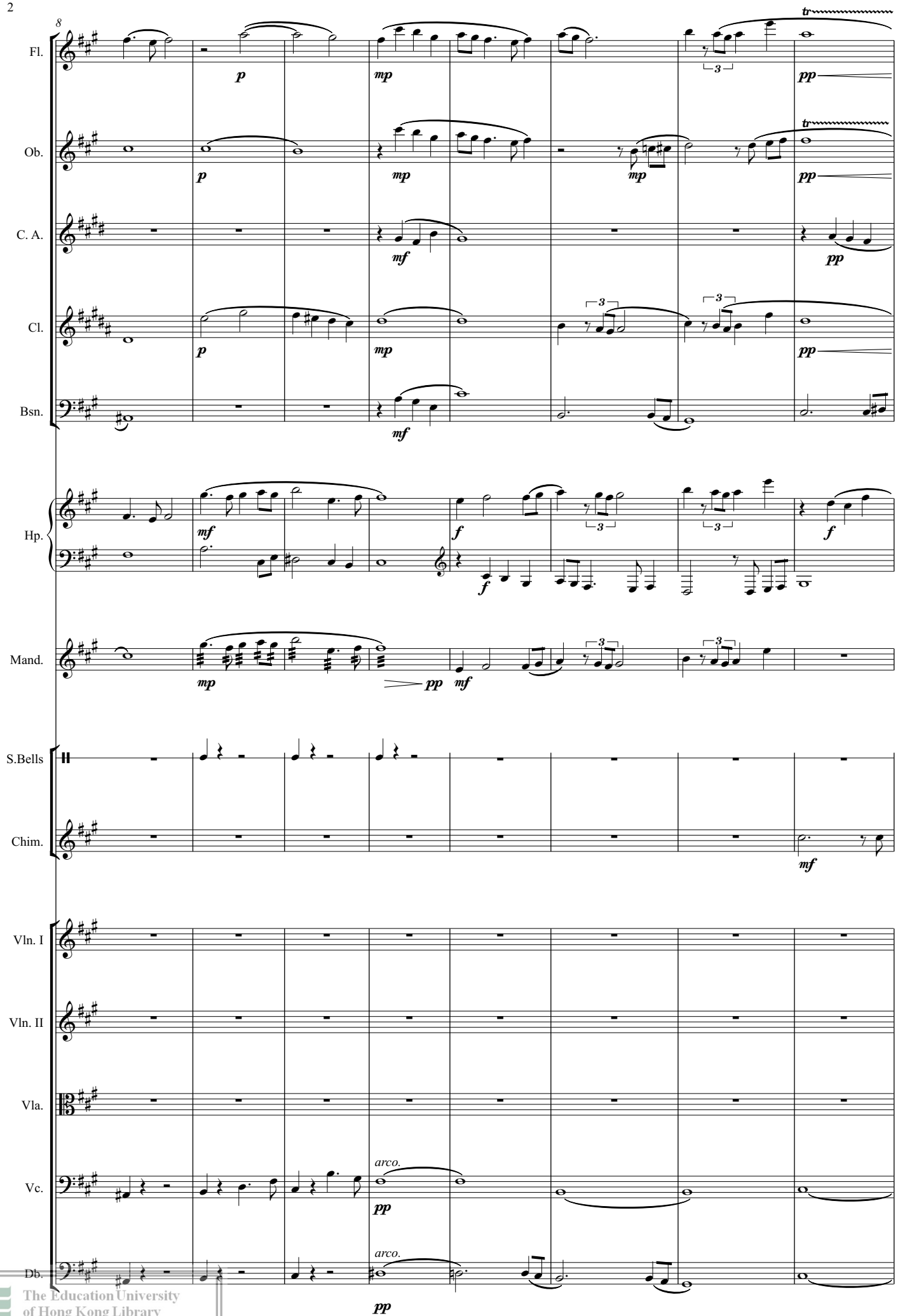


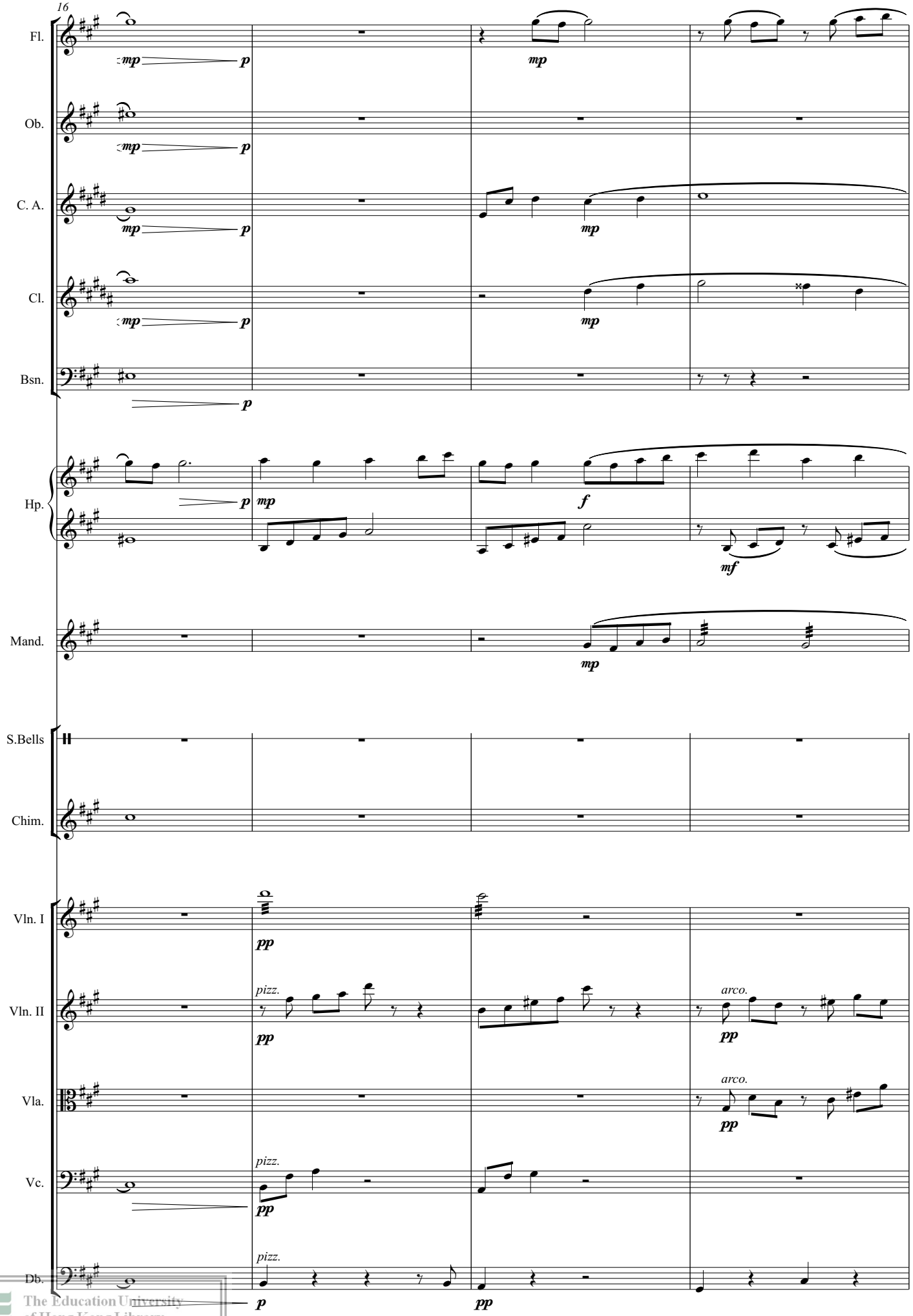


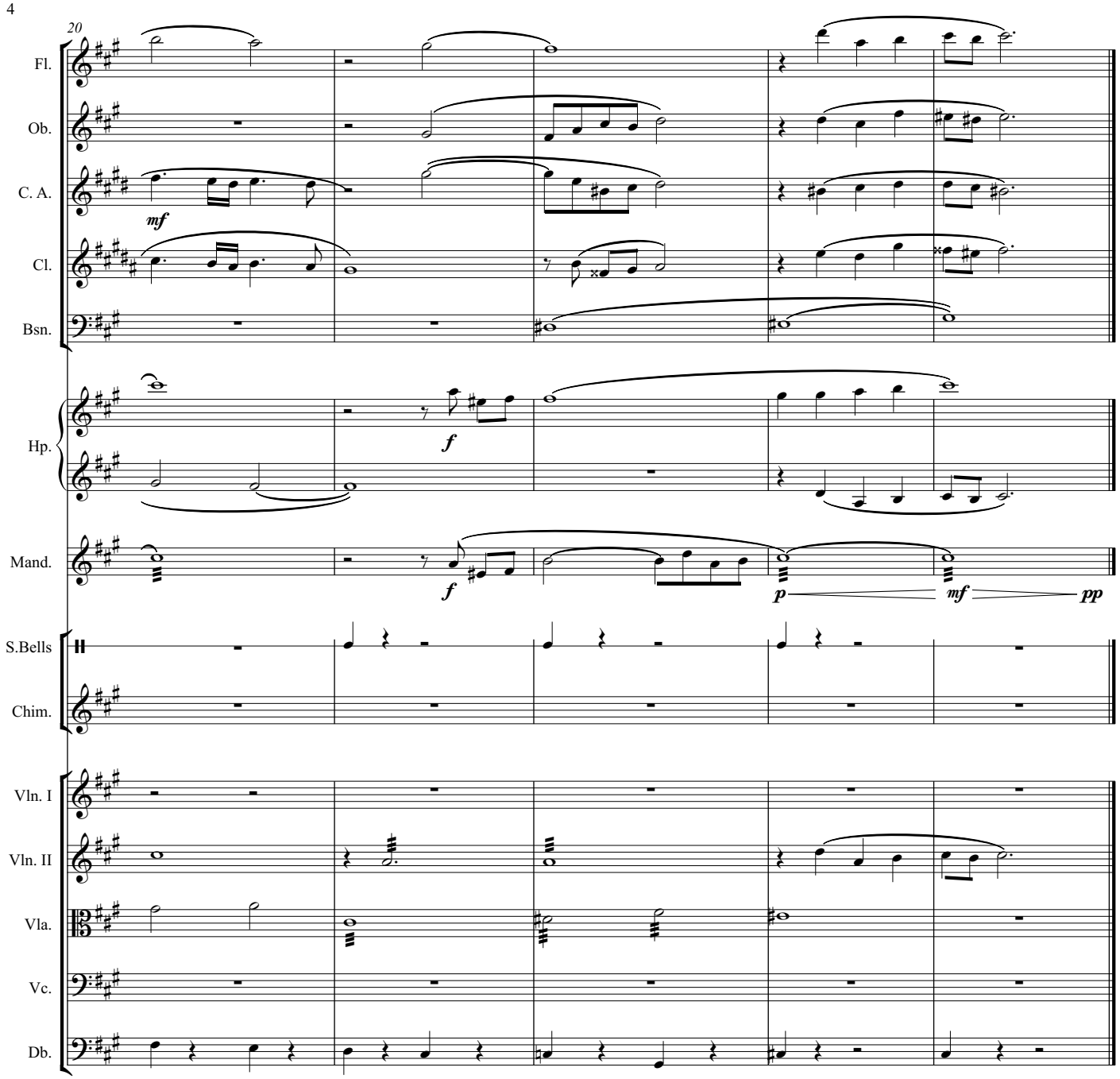
Appendix II

Astera's Parhelion - Full score (Player's hideout) - Shortened version



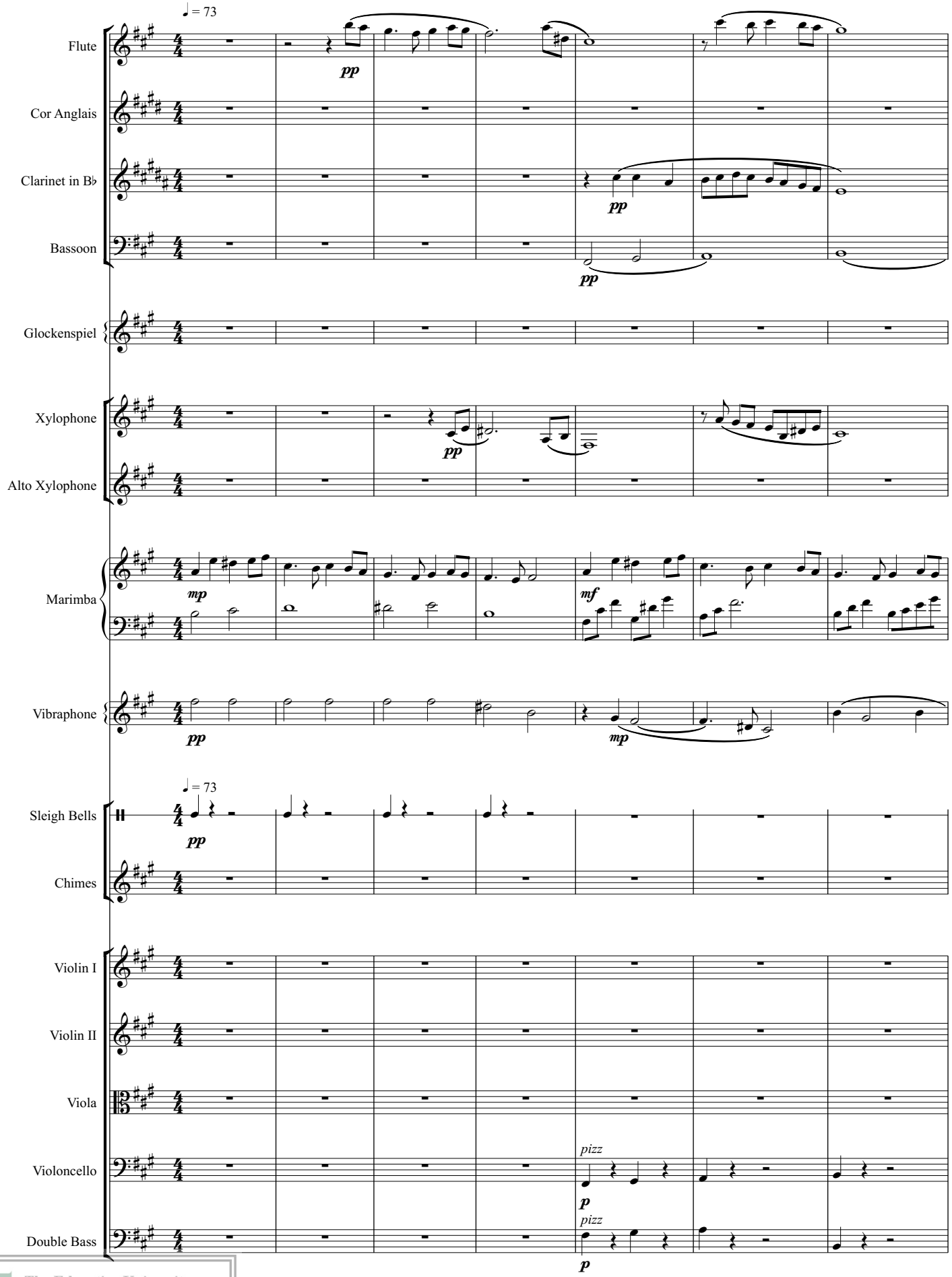


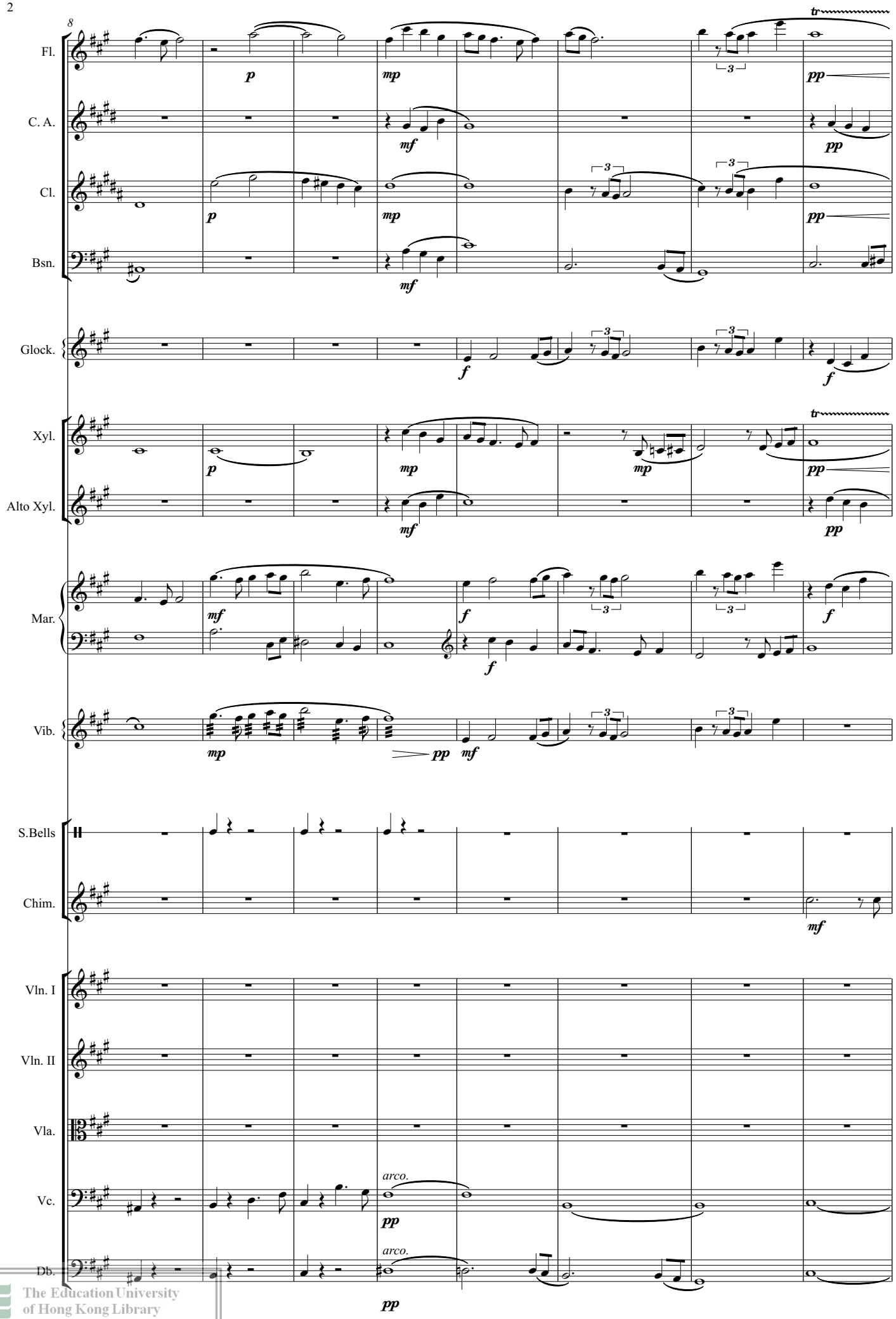


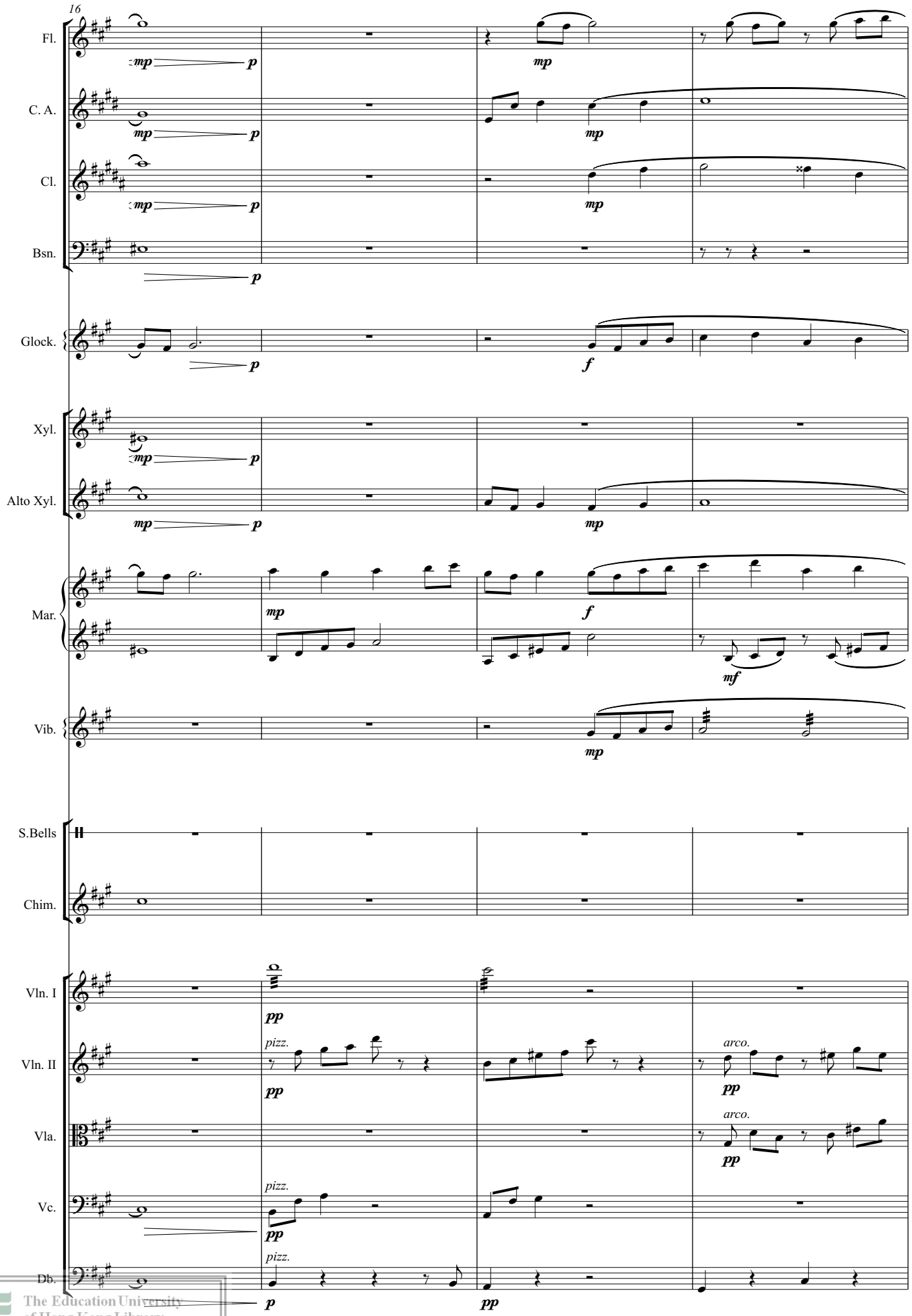


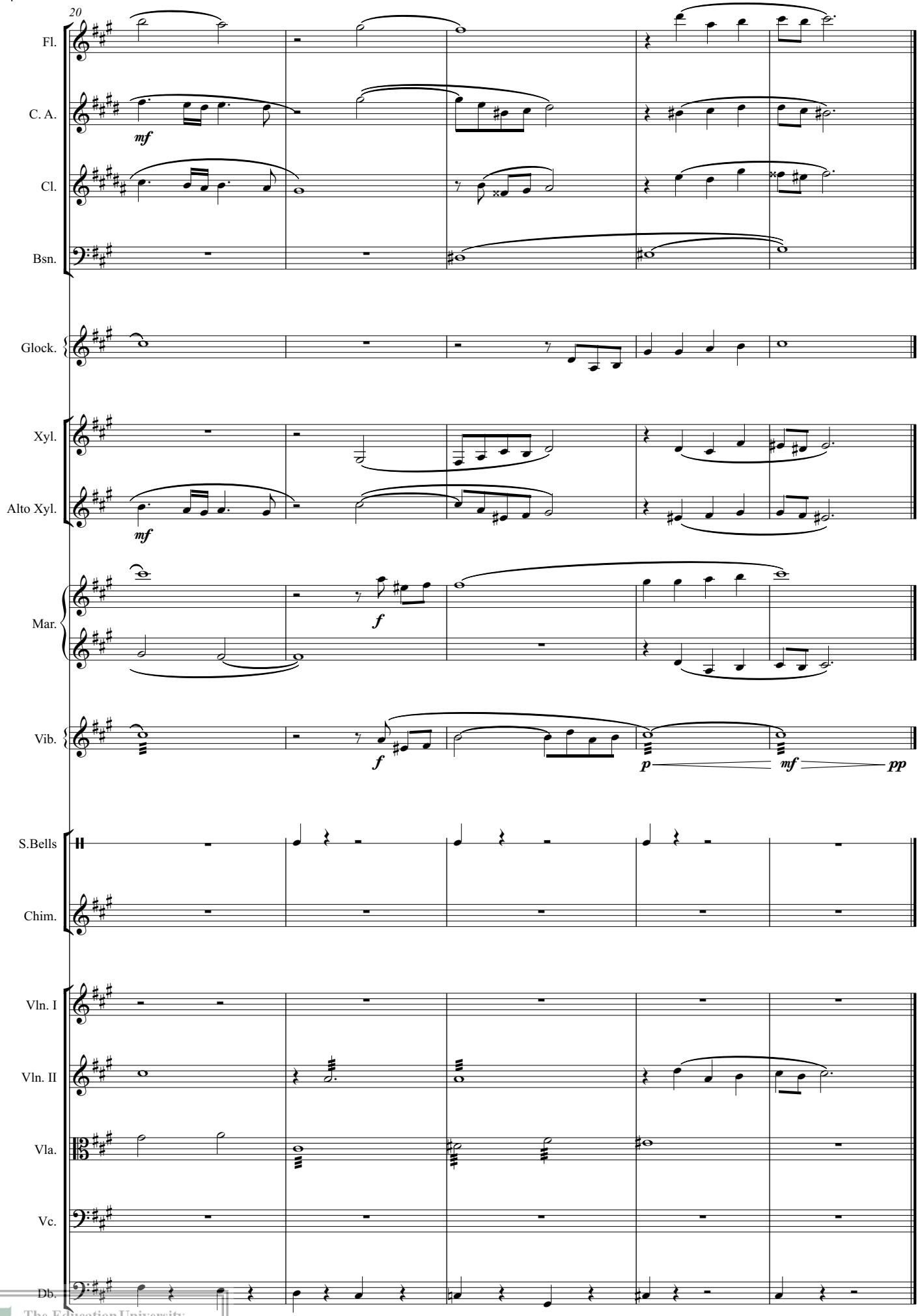
Appendix III

Astera's Parhelion - Full score (Research base) - Shortened version



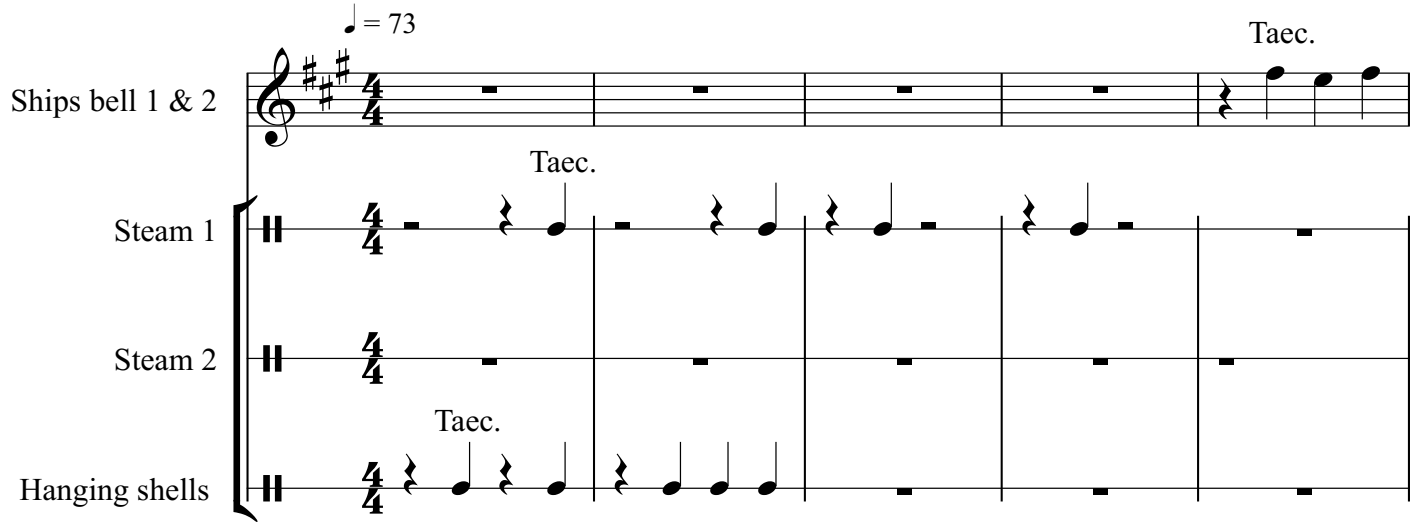


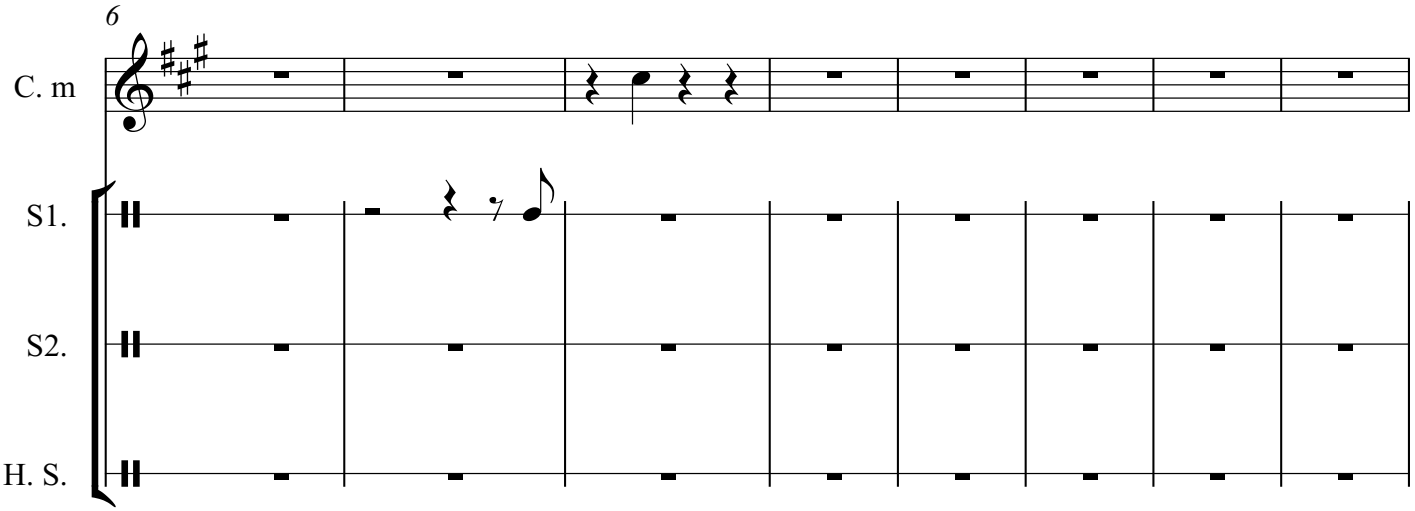


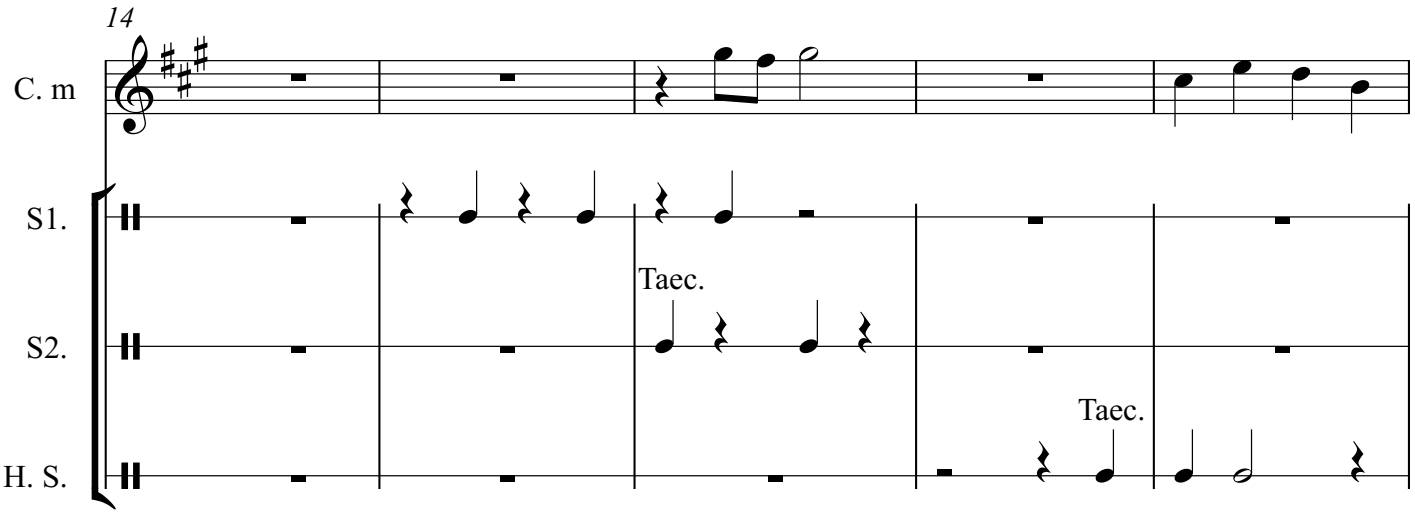


Diegetic reference - Research base

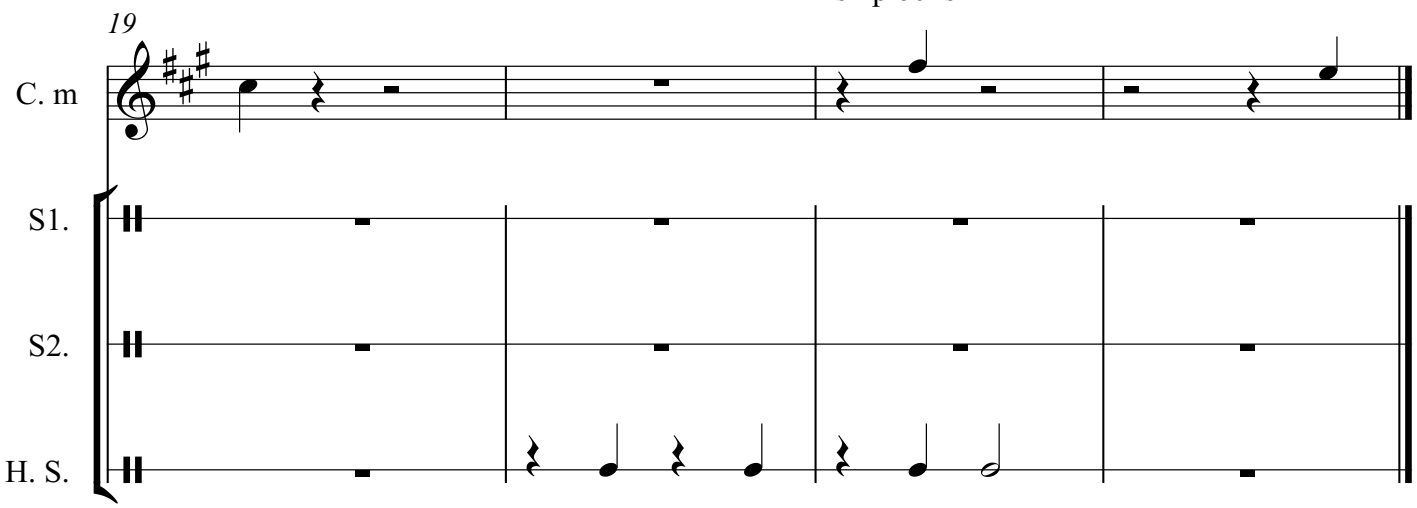
Triggers accodring to event cycle = Taec.





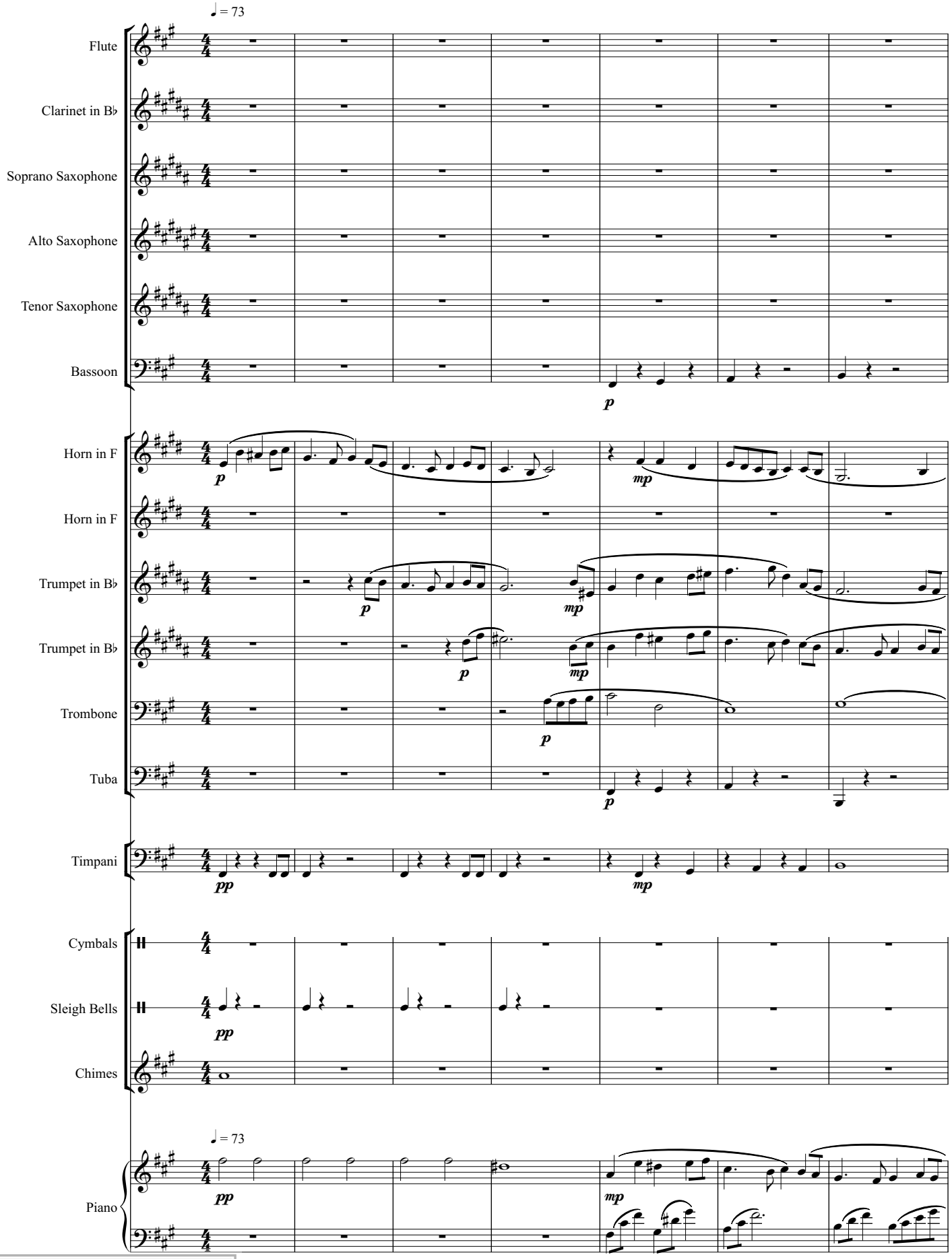


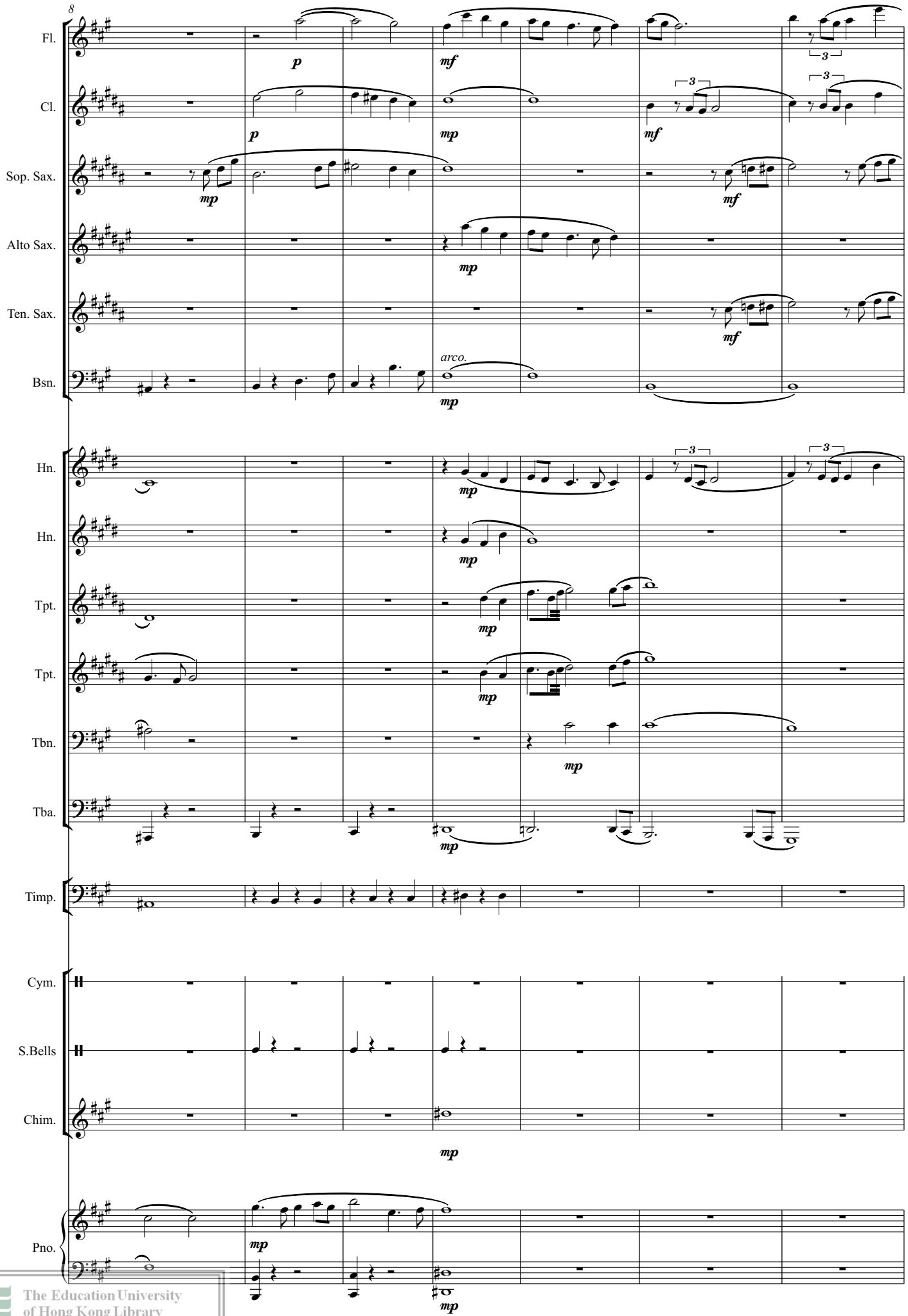
ship bells 2

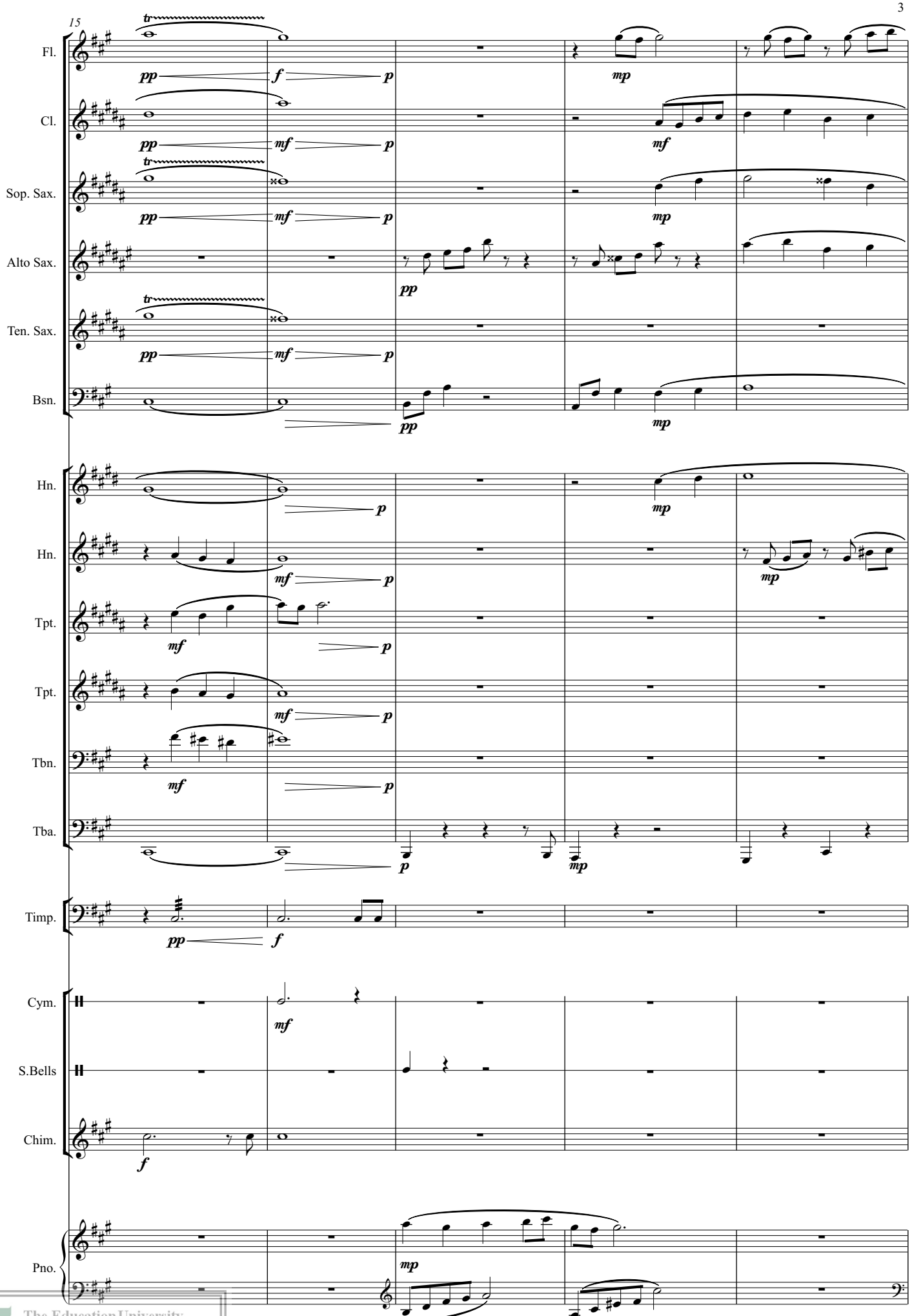


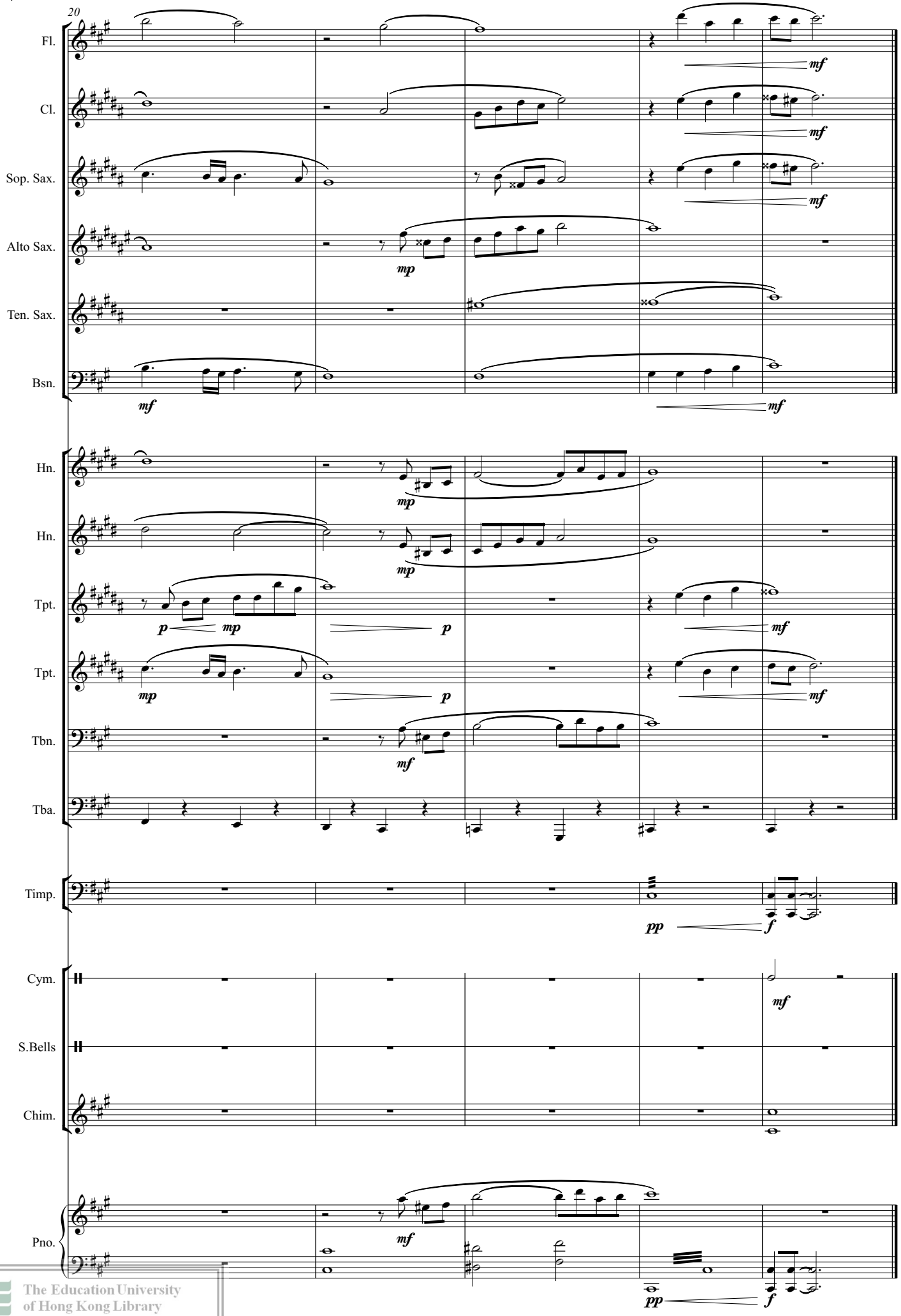
Appendix IV

Astera's Parhelion - Full score (Workshop) - Shortened version









Diegetic reference - Workshop

Triggers accodring to event cycle = taec.