Title	Piece : a bases for generating an integrable music video annotation search system
Sub Title	
Author	Mannschreck, Ryan Nicole Kunze, Kai
Publisher	慶應義塾大学大学院メディアデザイン研究科
Publication year	2018
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2018年度メディアデザイン学 第655号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40001001-00002018-0655

Master's Thesis
Academic Year 2018

Piece: A Bases for Generating an Integrable Music Video Annotation Search System

Graduate School of Media Design,
Keio University

Ryan Nicole Mannschreck

A Master's Thesis
submitted to Graduate School of Media Design, Keio University
in partial fulfillment of the requirements for the degree of
MASTER of Media Design

Ryan Nicole Mannschreck

Thesis Committee:

Associate Professor Kai Kunze	(Supervisor)
Professor Matthew Waldman	(Co-Supervisor)
Professor Naohito Okude	(Member)

Abstract of Master's Thesis of Academic Year 2018

Piece: A Bases for Generating an Integrable Music Video Annotation Search System

Category: Design

Summary

Picture in your mind's eye a music video that intrigued you. Whether it was the vibrant colors, the timely choreography set to the rhythmic beat, or the pretty face emoting, place that into your mind. You want to see it again. Enjoy those five minutes one more time. What was the name of the song? Can you remember the artist, any of the lyrics? No, the only thing that lingers is the images and the humming of the melody. How will you find this nugget of media in the large dark mine of the Internet? Music videos are characterized by their marrying of visual content and musical content. Despite music video's existence as a dual component art form, they are left to only be classified by their lyrical component. Those seeking music videos work with a limited tool set to find their chosen content, such as searching by the metadata of the song. This method completely ignores the visual aspects of the media. Welcome Piece, a new augmentation system created as a method for searching music videos by adding to existing music recommendation systems. The aim of Piece is to contribute to the initial development of an augmentation to pre-existing music recommendation systems. Extending the systems range with the addition of a music video search ability by integrating tags inspired by the visual elements of videos. Addressing cross-cultural/multilingual difficulties. Allowing for new ways to share, enjoy, and use music videos. First focusing on the music videos produced in the Korean pop music industry. The thesis concludes with the acknowledgment that further study is needed by branching into other music genre's, finalization of Piece's system, and full implementation of Piece's ability to be integrated into existing systems.

Keywords:

Annotation, Tagging, Korean Pop, Music Analysis, Video Analysis, Music Video

Graduate School of Media Design, Keio University

Ryan Nicole Mannschreck

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Picture in your mind's eye a music video that intrigued you. Whether it was the vibrant color scheme, the timely choreography set to the rhythmic beat, or the pretty little face emoting on the screen place that into your mind. You want to see it again. Make it fresh in your memory again. Enjoy those five minutes one more time. What was the name of the song? Can you remember the artist, any of the lyrics? No the only thing that lingers is the images and the humming of the melody. How will you find this nugget of media in the large dark mine of the Internet? That is where Piece comes in.

Music videos are characterized by their marrying of visual and musical content. Despite music video's existence as a dual component art form, they are simply categorized by their musical qualities. Left to be classified by the genre given to its lyrical counter part. Those seeking music videos work with a limited tool set to find their chosen content, such as searching by the metadata of the song. This method completely ignores the visual aspects of the media. The aim of this thesis is to contribute to the initial development of an augmentation to pre-existing music recommendation systems. Extending the systems range with the addition of a music video search ability by integrating tags inspired by the visual elements of videos. Giving the users the ability to create their own tags and uses. Presenting Piece, a tool music search systems can add to their existing programs. Piece endows the system the ability to search for music videos by their visual qualities, metadata, and by user suggested tags. No longer will you have to scramble in the dark. As long as you have a piece of the puzzle, Piece will give

you a peaceful search experience.

1.1. Proposal

Piece is a multi-faceted tool developed from four fields of data. We combined existing genre categories from music, movies, and television, company produced metadata, user generated metadata, with current music recommendation systems analysis's to generate the bases for our new hybrid system that fits the dual nature of music videos. This would require formulating tags to facilitate the annotations. Tags are "unique identifiers for a component or line segment found in the data cache not expressed in text but in data." [2] Annotation though similar is not the same. Annotation is "information about a component or line segment that appears on a drawing. This text can include the tag property of a component or line segment." [2] Piece is built on the principle of utilizing users natural language to create these tags. Piece can be used for self-expression; greater ability for categorization, detail oriented sorting, and quick searching. Therefore understanding the people who use it leads to a more user friendly interface based in the words and phrases they themselves would use. Through annotation, categorization, and organization a platform for searching music videos can be created. Piece builds something new from the foundations of what already exists just like music videos themselves. To reiterate our new platform, Piece, is a multicultural music and visual based system that has room for growth. It allows for professional and personal categorization. It begins addressing cross-cultural limitations of old systems. It is not a finished product but rather an exploration into a new protocol for system development.

1.2. Parameters

This thesis is a grounded theory study. Beginning with an initial exploratory observational research study to look into related works and current annotation systems. Followed by an observational research study done on the common search and consumption habits of music video's consumers. Using these exploratory searches as a jumping off point, the next stage was that of a problem orientated

qualitative research method using cross sectional studies consisting of descriptive surveys and interviews before and after witnessing the music video samples. The first cross-sectional study was done with a small sample with a diverse cultural background.

The perfect system for organizing and recommending music is a commonly sought out dream. Tagging music in itself is difficult. From troubles in cross cultural categorization, language barriers, and an industry that cannot agree what qualities constitute a genre, such a goal seems futile. Different professionals might have differing opinions between what qualifies as say an indie rock song or a pop punk anthem. [7, 33] Companies and studies have worked several angles in an attempt to solve these problems. All of these focus solely on annotating music. This thesis looks to apply their techniques to another medium of the music industry. Music videos require one to analyze both musical content and visual content.

Taking into account the troubles found in the music industry, sorting genres in their culture and then attempting to globalize them, looking into a specific cultures genre seemed prevalent for the initial study of Piece. Turning to K-pop was a choice geared toward addressing these concerns in the industry. Therefore, this thesis primarily focuses on the music videos produced in the Korean pop music industry. This genre of music was chosen due to the fact that Korean Pop music (Kpop) is an industry heavily reliant on music videos to distribute its content and bring in new consumers. Beginning in 1999, hallyu (the concept of Korean popular culture) has been steadily growing. Showing its international rise in 2009 and breaking into a 5 billion dollar industry as of 2017. [11] It has a large global fan base despite the language/cultural barriers. According to the Korean Foundation with aid from the Korean Foreign Ministry, as of 2015 International fans reached 35.59 million, an increase of 63 percent since 2014, at 21.82 million. Kpop fan clubs are now in over 86 countries roughly about 1,493 in total. [10] The totals of 2017 have currently not been translated into English. Music videos are important to the industry due to their ability to be the gateway for many international fans into Kpop. There is a large variety of styles but visual aesthetic is a major factor in all Kpop videos. The most viewed Kpop song as of 2018 is Psys Gangnam Style at 3,038,872,419 views on YouTube alone. The top viewed K-pop video on Youtube in 2018 is currently "TT" by Twice with over 300 million views on

YouTube. The average views on YouTube for Korean pop music videos tend to be closer to 200,000 views. [37] These totals do not reflect the all of the views from every site on the Internet of each song. If looking at the Korean viewing base one has to look at other music video websites totals such as Bugs Music, Naver Music, M.Net, and Melon to name a few. Countries such as China and Thailand also have their own preferred websites and apps to access music video content such as, JOOX or COOLISM. Korean pop music stands as a testament to the power of music videos for crossing cultural lines.

For some music consumers K-pop means music made by idols. Idol is a term commonly used to refer to musical performers who are promoted as total packages, dancing, singing, good looking, and more about this manufactured product then the music itself. As for Koreans, K-pop can mean anything from Idol music, to music with upbeat tempo melodies, and/or music made by large companies. They have a separation for their genres of music. Where as internationally music produced in Korea is often placed under the umbrella term of Kpop. [24] So how can these distinctions that happen in Korea be made and understood overseas? Especially if even industry professionals have a hard time agreeing with what constitutes each genre in their own country? This problem reflects the trouble of the music sorting industry at large.

The pre-study and the tag identification study where conducted in English. The tag identification study was also done in Japanese. This project is in the process of being conducted in the Korean language. The reason for choosing English is that currently most global music companies and music distribution sites use the English/American music genres system. English made sense as the beginning language for identifying words and phrases for tagging the music videos. Japanese was chosen due to current access to the Japanese population and it's large consumption of Korean pop music. Both languages allowed for testing in the language that is not utilized in Korean pop music. Sometimes Korean pop songs are translated into Japanese. Some are also created solely for Japanese distribution. The only other languages that sometimes get this treatment is Chinese and English. This was a marker of their high rate of consumption of the K-pop industry. This study was comprised of a larger sample. Diversity in cultural background, age, and gender was factors looked for in sample selection. This thesis

used Kpop music videos as its sample base for content.

1.3. Contributions

- 1:Insights into how and why people search for music videos.
- 2:Differences in tagging and search practices based on culture/background.
- 3:Generation of annotations from reported responses individual to user's specific backgrounds. Using posture and facial expression (based on active and passive posture) as guides to annotation placement in content.
- 4:Possibility of auto-generation of specific tags for current and future content.

1.4. Thesis Structure

The thesis is as follows. Chapter two will review related works and companies in the field of music organization and/or search engines. The chapter will break down their approaches, positive attributes, and short comings. It will finish by summarizing the total of available methods and how this lead to the creation of Piece. Chapter three presents the design process undertaken to develop the future structure of Piece. Chapter four looks over the evaluation of the tag identification studies conducted for Piece as a means of creating a bases for its data structure. The chapter will describe the guidelines on how to implement Piece's structure. Chapter five will look at the study conducted to show proof of concept in using natural language as a form of tag generation. Chapter six discusses the ongoing collaboration with professor Nam and his students at KIAST. This thesis ends on chapter seven giving the conclusions, limitations, and possible future works that came as an outcome of the research conducted.

Chapter 2

Related Works

Music videos arguably came about in or around the 1920s with the invention of soundies, talkies, and musical short films. More commonly it is believed that music videos, as we know them today, came about in the 1980s. The art form being made popular by music video shows such as Countdown and MTV. Since that time music videos have been used as an artistic form of expression and/or a promotional device. Music videos have been used for all genres of music and in multiple countries music industries. Despite their long-standing role in the music industry they are often left second-class to their song counterparts. Music videos are rarely categorized separately from their song genre making looking into specific content, moods, and or styles difficult. Thats not truly recognizing the visual effort put into the cinematography required to develop a music video. The styles and content that they create are important contributions to the music industry. There are companies and industries specifically devoted to putting these visual expressions to their musical counterparts. Despite the money, time, and effort pored into these creations they are not given recognition. When searching for music videos you cannot look for them by their visual media. One has to search by metadata alone. What if you want to find something specifically based on a color palette, a mood that it gives off, or a story that it conveys? You cannot find this by the name of the artist or the genre. That is where this thesis begins.

In this the digital age access to music is substantial. The need for greater organization and retrieval is imperative. Users are no longer restricted to a physical storage space for their music collections. Consumers have access to hundreds of

thousands of songs and the number is growing everyday. Organizing and searching these vast databases is a feat. Existing systems function but are far from perfect. Relying on outdated classification sets, limited information, and globally ignorant categorizations. Meta data, such as song name, album title, etc. is a common tool used but it is time consuming labor-intensive work, not to mention expensive. [31] It requires a large scale content analysis and web-mining of all content wanted in the system. It can only be done by paid professionals, often within the company. It is also limited. Meta data is just the facts about a song such as, artist-name or album-name. It does not address any of the intrinsic musical or vocal elements of a song beyond the genre classification. Genres, often found in the metadata, tries to classify this information. Humans have the ability to process, recognize, and analyze sound based on its construct. Humans can build connections between songs and their elements distinguishing patterns. Do to this ability humans can categorize music into genres. [31] Genres are a frequently used form of organization for artistic media. Commonly placed as a part of the songs metadata set. In Music Information Retrieval (MIR) research a great emphasis on genre is placed when it comes to organizing and retrieving musical content. Genre is often designated through feature extraction of the music. Feature extraction is when a segment of audio is analyzed and characterized. This characterization is often expressed in a compact numerical representation. This segment is built from analyzing elements such as pitch, timbrel features, and rhythm. [42] Genre schemes are commonly used to structure collections of musical content. Genres are not infallible however. They are far from objective and are inconsistent. [1, 29, 40] They are highly dependent on the listeners behavior. McEnnis and Cunningham explain that "music is interpreted in terms of how it expresses local issues and concerns." Do to this the understanding of the music can be radically changed from its intended inspirations or meanings." [28] The more distant or removed the culture from the producing culture the greater the potential for distortion. Studies pertaining to music genres and classification schemes concluded that they are nether objective, consistent, or reliable. There is not a strong conscious on what marks a genre. [9, 35, 42] Despite these grievous shortcomings genres still remain a comfortable way for users to search for their music. It still functions as a primary search quality. Provided access via bibliographic info and genre scheme

assumes the users prior musical knowledge for music retrieval.

Generally speaking most music recommendation systems existing today use user-supplied metadata, web-based metadata, and company given metadata to supplement the old method of content-based systems. User-supplied metadata is still developing. Currently the most sought after system is one that incorporates audio characteristics into its framework. Individuals have their own organization systems. An example being, one music listener organizes their collection by music genre while another might organize it in alphabetical order by artist. Humans build an idea of music analysis that is fully personal. In Bennetts Popular Music and Culture, he sums up the process in which consumers add layers of meaning to the musical content they consume. "Consumers take the structures of meaning the musical and extra musical resources associated with particular genres of pop and combine them meanings of their own to produce distinctive patterns of consumption and stylistic expression." [4] The idea for Piece bloomed from this circular relationship between humans and music.

2.1. Existing Systems

A music video annotation system does not exist. Search systems for music videos such as, YouTube, do exist but it does not have the ability to find videos based on content tags but rather on metadata tags only. Music recommendation systems are commonly based on two approaches to sorting: content based analysis and collaborative filtering. Content based analysis looks at the lyrics, vocal qualities, and musical make up of a song to generate tags that can be used to find the song. Content based analysis organizes its database by these intrinsic features in a song. [28] Examples of experimental systems using the content based analysis strategy include, Pampalk et al.'s [34] system with the exception of user's abilities to augment the system with explicit tags, Logan's [27] generating recommendations from similarity, and the systems proposed by Chen and Chen. [6] So if a person is using this system they will be able to find songs similar to say the harmony, RPMs, melody, etc. Pandora.com is a commercial example of this type of system. Collaborative filtering looks at the relationship between users and musical items to make predictions. It closely follows user's behavior and responses

to build how it will recommend music to the user. A user using this system would be able to search for music based off of statements like, 'If you like... then you might like...'. Multiple commercial services use this strategy such as, ITunes, Spotify.com, and YouTube.com. More and more frequently music recommendation systems have begun incorporating user-supplied meta-data to their pre-existing content-based approaches. It allows for cheaper information gathering as well as user customization. Users would be able to search for content by more opinionated terms such as moods and/or user curated play lists. Spotify.com and 8Tracks.com are examples of this system form.

2.1.1 Commercial Systems

Spotify

Spotify is the only on-demand music streaming service that is not a web-based service. [19] Rather, it uses a peer-to-peer network (p2p). 8.8 percent of the music playback comes from Spotifys servers. The peer-to-peer network handles 35.8 percent. Spotify also employs the user's local cache at about 55.4 percent. Spotify on smart-phones is the only part of their system that gets all the music directly from the Spotify servers. 61 percent of playbacks are done in a predictable order. This means listening to an album front to back. Spotify pre-fetches the full album or playlist so that it can play instantly. Not only does Spotify queue up the album it queue up subsequent tracks. Only 39 percent of playback is accessed randomly. This is when users cherry pick through tracks rather then full albums or lists. Another unique quality of Spotify is the users ability to share playlists and tracks through social media. Users can link their Facebook or other social media accounts and publish what they are listening to or share directly with other users. [19] Unlike Pandora.com Spotify does not build a playlist for you from the kinds of music you like. That said a version of this, Spotify's "Artist Radio", is in current development. Spotify builds its recommendation system by user habits. Collaborative filtering is the primary in their search hierarchy. Layer two comes from direct and indirect user-supplied metadata mixed into their pre-existing content-based algorithm. User supplied data comes from tags given to songs and playlists by users. Users are able to curate playlists. If they want to

make them searchable they have to add tags. Giving simple hash tags such as, happy, sad, sexy, or party. This can also get a bit more outlandish. Some users tag playlists with tags of anime characters, celebrities, or literary characters they feel the playlist fits. This displays some of the flexibility of humans categorization methods. An ideal system for Piece to be added to is Spotify. One, it has no music video function whatsoever. Spotify only works with tracks. There is no conflicting or pre-existing system to get in the way. Two, its search system already works in a kin to Piece's. Working on a multi-level hierarchy, collaborative filtering, user-supplied metadata, tags/annotations, Spotify only misses three of Piece's infrastructure. Vocal annotation, musical annotation, and international search assistance. Of course it also misses Piece's keystone, music videos.

YouTube

YouTube is a juggernaut online service when it comes to viewing videos. This includes music videos. Record labels around the world have their own YouTube channels in an effort to get hype from it's strong user base. YouTube employs a simple but effective algorithm for calculating relevancy of videos to searches on its website. First and foremost is the metadata. The metadata has to be added by the user who uploads the video, YouTube does not do this itself. The metadata can include titles, artist name, company, etc. Under the label of metadata users can also add tags and the lyrics/captions to increase search ability. Secondly comes popularity. How often is a video searched for and viewed. This part is recorded and kept by YouTube. [14] This method is imperfect. It does not count individual people who view the content but the total number of views. YouTube also employs a thumbs up and thumbs down ranking. Users can ether thumb up or down videos, the more ups the higher the ranking of the video and vice versa. To augment some of the troubles this causes YouTube also counts watch time. How long do users stay and watch the content? Do users watch the video in full? YouTube does not have a robust MIR system in place. Despite the fact that YouTube mainly gets its information by user generated metadata it does not allow for much customization. It's recommendations can be repetitive, loosely related, and even sometimes feel random to users. Recommendations usually come from videos with similar keywords in their title or by a 'viewers that watched that

watched this' style. [14] This means that users are limited to presumed knowledge of musical information. If a user does not know the bibliographical information of the desired video they are left with little options.

Since systems like these are already in place building a whole new product is ill advised. MIR research and the companies that implement their findings are constantly updating their systems and tactics. Piece comes from an understanding of this industry landscape.

2.1.2 Experimental Systems and Cross-cultural Music Information Seeking

The amount of research done by in MIR on non-western music is minimal at best. Futelle and Downie [8,13] bring to light this gregarious oversight in the MIR research base. Whether it is a Westerner searching for none western music or vise versa all people who seek music information are plighted by the incapabilities of current systems. The largest hurtle in this arena is language. Metadata often comes in the language of the contents origin. If a user searching is not native in this language they face navigation issues. What if this user cannot remember the artist, title, or album? What if the information is in a different writing system? The Roman alphabet doesn't work for writing Korean or Chinese. It requires translation and not all companies are willing to do this. Independent artists can not always afford to do this. If a Korean user wants to find a Japanese song, how do they go about it? Lee, J. H. [26] looked into how Korean users of Google Answers and Naver, a popular Korean 'knowledge search' portal, navigated these challenges through group information mining. What they found was a reliant way of finding answers to vague, outlandish, and incomplete quarries about music. Most questions where not only answered but answered correctly and verified. Quarries often consisted of sections of lyrics, descriptions of visuals, or desired traits/musical qualities. These traits could be asking for songs appropriate for special occasions (weddings, birthdays, parties), moods (upbeat, good for studying, meditation), or for specific people (father, sister, etc.). [26] When asking after their quaries users could remember visual characteristics. Sometimes key components of the searches rested in related or 'associative metadata' such as the song

appearing in a movie or TV show. Not useful for an MIR system but it is if the MIR system is used for music videos.

2.2. Insights

One tactic to be undertaken in the development of Piece might be to try an identify tags that are culturally neutral. Tags that are multilingual or could be used by multiple cultures. In McEnnis and Cunningham's exploration on the contributions sociology can bring to music recommendation systems they recommend a shift in the definitions of sub-cultures and neo-tribes when applied to construction of a multicultural system. Sub-cultures are described as "a social group that is significantly different from others and requires a significant commitment to join." While as a neo-tribe refers to "a loose group of people that share a subset of musical tastes." They state that a system that can recognize the differences between these two groups would go a long way in it's prediction capabilities. [28] Lending the ability to distinguish rigidly defined personal definitions of a genre or style of music from users who stay on the fringe of the groups musical contributions. In an attempt to develop a multicultural system Piece will have to understand this thin line and execute it when performing it's duties. Lee's study located seven main queries pertaining to music on Naver and Google Answers, "Identify artist/work, get recommendations, acquire lyrics, request translation, locate specific version of work, seek information, request transliteration, and to locate a work." [26] The percentage at which these queries where more commonly asked showed significant cultural differences. This significance opens the way to classifying those features that could be considered universal ways of inquiring verses those that are more 'culturally determinate'. When composing the surveys necessary to begin gathering tags for Piece keeping these seven types of inquires is wise. Focusing on how it can meet these needs or render them mute. For Piece a system that has facilities like Bainbrige's et al [3] a 'query-by-example model' for the MIR system might help alleviate some of the strain caused by users' wide range of descriptions and search terms.

Chapter 3

Design

Piece is a concept born of a deceptively simple dream, for a music video recommendation system. Piece is a tool developed to augment MIR systems or work as an independent music video search system. Piece by combining existing genre categories from music, movies, and television, company produced metadata, user generated metadata, with current music recommendation systems analysis's generates accumulates into a new hybrid system that fits the dual nature of music videos. Piece formulates tags to facilitate the annotations necessary to run such an undertaking. The tags where gathered from users, lyrics, vocal annotation, and pre-existing systems to form a pool of data. These sources where organized into hierarchal layers to build a top-down system of categorization. Layer one saw that each music video from the sample was categorized by combining music, movie, and TV genres to create holistic genres that music videos can be placed under giving credit to its dual media style. Layer two, incorporated existing annotation systems used by 8tracks, YouTube, and Spotify. One facet that makes this annotation system somewhat unique is that it takes into account the commonalities and differences between cultures and languages when consuming musical content. It goes directly to the people and asks them what terminology they would use to describe the content they just consumed. Though its greatest uniqueness comes from what its being applied to, music videos. Layer three recorded what people remember, look for, and enjoy in music videos. What is it about the visual portion of music videos that draws people to want to watch them? To begin addressing cross-cultural needs we conducted layer three in two languages, English

and Japanese. A form of this study is currently underway in Korean. The answers provided by the participants in each language would then be combined and translated into each language allowing for an expanse of tagging terminology for each culture. Piece is ultimately an augmentation tool. An extension to the MIR system geared toward the integration of music videos into the commercial world of music recommendation systems.

3.1. Design Process

Piece's design is a guideline for implementation. It is not a full fledged product, but rather an exploratory work to be take as a guideline for building such a system. Piece currently is a database pool with instructions. Figure 3.13.1 shows the five areas of data Piece pulls from, existing annotations, company meta data, user-generated tags, professionally-generated tags, and global genres. The first data to be added into the Piece system was the metadata. Title of the song, artist, composer, year of the song, album name, etc. All of this data can be pulled from the online databases of the companies owning the rights to each of the songs. Some of the metadata from the music video itself was added into the system. This includes but is not limited to the director of the music video, the stylists, make up artists, the video company, and creative directors of each video. The lyrics of each song was imported and incorporated into the tagging system's database of words, phrases, and sentences. The lyrics where imported as both a whole and as key adjectives and nouns found within the lyrics. For any videos that had any additional dialogue that was included in the music video but not in the song itself, the script was also imported for these dialog sections alone.

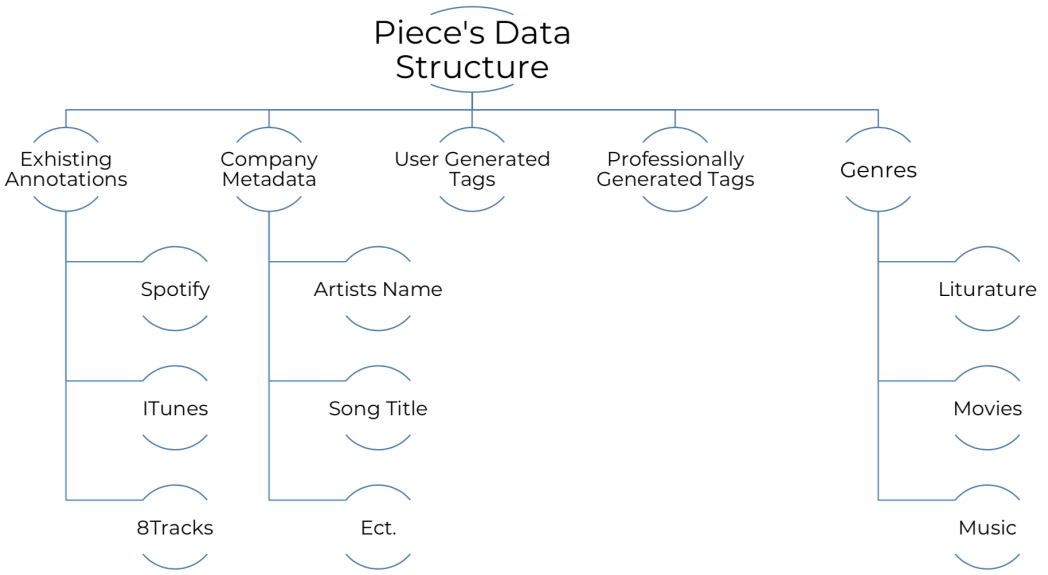


Figure 3.1: Demonstration of Piece's data gathering resource structure.

It was important in the construction of Piece to address the user's capabilities for influencing the development of the system. Tags should reflect the way in which users speak, organize, and process information. Tags generated by our research and by Piece needed to be derived from the people. They also had to be malleable. Capable of changing along side the evolution of language. It was determined that all studies would focus on the users moving forward.

3.2. Pre-Study

3.2.1 Methods

The first exploratory study needed to verify if indeed people retained information based on the visual portion of music videos. It also needed to see if Kpop videos where a viable genre for the study. Finally it had to begin shaping what information would need to be gathered from later participants for the tagging system and what was superfluous. For these reasons a cross-sectional study was

chosen. By surveying the users after each video and at the end the potential use of particular questions and information could be assessed for worth. This exploratory study would form the bases for the primary experiment that would lead to the feeding of the potential Universal search system.

3.2.2 Music Video Selection

Step one consisted of selecting videos that would be used in the initial study. This required creating an intricate selection process, as it would not be possible to show an unlimited number of videos. The videos selected had to encompass a wide range of visual and musical styles, displaying the types that could be found in Kpop videos. The algorithm created for this purpose looked at the total number of views, the genre of the song, the style of the music video, the popularity of said video, the company that made the video, the company that produced the video, the gender of the artists, and whether it was a group or solo artist. Videos had to be no more then five minutes long so as not to press our participants ability to concentrate.

Total number of views was spread out between highly viewed videos and videos with a lower or more average total number of views. If the videos only showed popular videos it would constitute as a bias in the experimental conditions. The same could be said for using only videos with a low number of views. The decision was made to have a pool of ten videos with which participants would receive a randomized set of four videos to view. Ten was thought to be an adequate number to show the wide range of video/musical styles. In the pool of ten videos two videos from each of the main genres of Kpop music were chosen. Two club anthems, two ballads, two hip-hop style, two sugar pop, and two electronic songs to create the total of ten songs to be used. All videos embraced a certain difference in style. Some using bright overly saturated colors others toned down and soft. Videos showed a range of mood and settings, some sad, some happy, some realistic, some fantasy. Some of the videos had story lines, others where more vague or artistic in stylization. Popularity of the videos was calculated from the number of views and their ranking on common music sites such as ITunes, Spotify, and M.Net. Any awards won for the video content was taken into account in this category as well.

There are four major Kpop music companies in South Korea: S.M. Enter-

tainment, YG Entertainment, JYP Entertainment, and FNC Entertainment, as well as smaller recored labels such as, Big Hit Entertainment and CUBE Entertainment. Representation from all of the major companies and some of the more prominent small companies was represented in the selection. Companies that created or directed the videos where taken into account for a diverse representation of the artistic styles of those in the industry. There is an even number of videos with male and female artists. There are not many groups that have mixed genders so it was decided to leave those off at the moment. An even divide between solo artists and groups, represented in both the male and female genders, was taken into account.

3.2.3 Participants

The initial study contained ten participants, two males eight females. Participants demonstrated a sampling of diverse cultural backgrounds. All participants could speak and understand English but it was not required to be their first language. All had little to no exposure to Korean pop music. None of the participants could understand, speak, or read Korean to any degree. The relevant information for each participant from the exploratory study is displayed in the table below. 4.1

Table 3.1: Participants

Name	**Age**	**Nationality**	**Gender**	**Sexuality**	**Languages**
P1	26	Italian	Female	Heterosexual	Italian, English, Spanish
P2	25	Japanese	Male	Heterosexual	Japanese, English, German
P3	23	Indian	Male	Heterosexual	Hindu, English, Japanese
P4	26	Thai	Female	Heterosexual	Thai, Japanese, English
P5	23	Chinese	Female	Bisexual	Chinese, Japanese, English
P6	26	Thai	Female	Heterosexual	Thai, English
P7	27	Chinese	Female	Heterosexual	Chinese, Japanese, English
P8	25	Chinese	Female	Heterosexual	Chinese, Japanese, English
P9	26	Italian	Female	Heterosexual	Italian, English
P10	27	Mexican	Female	Heterosexual	Spanish, Japanese, English

3.2.4 Procedure

Participants were asked to watch five Korean pop music videos. The music videos were selected from a pool of 10 videos, utilizing the Latin square method so as to ensure controlled randomization of the content. The videos watched by each participant and in what order is displayed in the figure below. 3.2 After watching each video participants were interviewed for qualitative data. Questions consisted of asking the participants for adjectives or phrases they would use to describe the music video, the movie genre they would give to the video, the music genre they would give to the video, what they believed the theme of the video was, and what they thought the key points of the video were. During the experiment, participants where videoed and the recordings saved. The videos where used to examine physiological responses by utilizing the open source program, Openpose, for analysis. To insure a baseline, participants where asked to sit for five minutes while recording their posture at rest. To make sure the recordings where in real-time a cover was placed over the camera before videoing and removed at the exact time that the video began with a simple system of a string that pulled the cover away from the camera lens once the space bar was pressed thus starting the video. Their eye movements where also tracked using the Tobii eye tracking system run through Hypermind to recorded visual interest. The Tobii eye-tracking system consists of a small bar placed at the bottom of the computer screen containing a series of sensors that track, capture, and record eye movements while looking at the screen. [16] It also makes note of when the eyes are at rest or are blinking. At the beginning of each participants session the Tobii system was calibrated to the participant's eyes using the inbuilt calibration system. Five minutes of recording was done to create a baseline of the participant's eyes at rest. Afterward an exit interview was conducted for more general thoughts on the music videos as a whole. A key part of the questioning was what part of the story they most remember and in which part they thought was the most important. This is significant not just because it registers their opinion but, because we can compare their responses to the physiological information we developed to see if there is a correlation of any significance. Did they show more physical reactions towards the same moments that they described as the most interesting? Did their eyes widen or did their smile increase when they viewed this particular section? Lastly participants where asked

questions pertaining to their consumption habits of music videos in their everyday life.

Name	Videos
P1	BTS-Fire, Orange Caramel-Catallena, Taemin-Move, Hyuna-Roll Deep, Block B-Her
P2	Luna-Free Somebody, BIGBANG-Fantastic Baby, BlackPink-Boombayah, Got7-Just Right, f(x)-4 Walls
P3	2NE1-I am the Best, EXO-Love Me Right, Luna-Free Somebody, Taemin-Move, f(x)- 4 Walls
P4	BTS-Fire, Orange Caramel-Catallena, EXO-Love Me Right, Hyuna-Roll Deep, Block B-Her
P5	BlackPink-Boombayah, Got 7-Just Right, 2NE1-I am the Best, BIGBANG-Fantastic Baby, EXO-Love Me Right

Figure 3.2: Videos Watched Sample

Recorded videos where run through Openpose. It is a free open software system that allows one to track the posture and facial expression changes of participants as they view the videos in real-time. The Openpose software works by utilizing the camera built into the computer. The software allows you to map key points on the body and face and follow their changes over time. [32] These key points are recorded as data, every frame is recorded and key points marked for that frame.3.3 The data recorded at each key point comes out as a series of numbers, marking the position of that key point. By calculating the difference between each changing point for each key frame one can see how the body and face changes over time. This allows one to map, for example, the upturning of the corners of the mouth that occurs when someone smiles. By matching these body posture and facial expression changes, it is possible to see what peoples' reactions are to particular moments in the music videos physiologically. [36] The data outputted by Openpose also comes along with a time stamp. This time stamp can be compared to the real-time video to see what part of the music video they are watching at the time of the change. Thus we gain a physiological understanding of each persons reactions while exposed to the content. This physiological data was compared to that of the responses each individual gave during his or her interview. Looking for a correlation between the memorable moments stated by the participant and

their posture. The data given from Openpose was compiled with the eye tracking data given by the Tobii system to comprise the profile of each participant.


```
{
  "version":1.1,
  "people":[
    {
      "pose_keypoints":[
        403.717,274.552,0.868576,407.634,345.011,0.578412,307.131,356.735,0.468358,0,0,0,0,0,0,508.096,338.486,0.445508,539.439,449.397,0.0697137,0,0,0,0,
      ],
      "face_keypoints":[
        355.192,253.364,0.49714,356.03,265.926,0.503663,356.449,278.488,0.499553,358.961,290.213,0.578728,362.73,301.938,0.678153,369.43,311.988,0.737321,
        .855,239.545,0.838864,396.23,242.476,0.880562,413.398,242.476,0.854861,422.192,240.382,0.814144,430.148,239.126,0.950248,438.942,241.639,0.819096,
        86.18,255.876,0.954179,379.48,255.457,0.957967,418.423,255.876,0.858732,423.448,252.107,0.983867,429.729,252.107,0.948434,435.173,255.039,0.84037
        2,390.786,299.426,0.882753,399.579,299.007,0.954697,402.929,299.426,0.953143,406.698,299.007,0.927054,415.911,299.844,0.862499,406.698,299.426,0.9
      ],
      "hand_left_keypoints":[
        0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
      ],
      "hand_right_keypoints":[
        0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
      ]
    }
  ]
}
```


Figure 3.3: Key Points of P4 from Frame 1286.

3.2.5 Results

Over all response to the content was positive. Four of the participants expressed a desire to continue seeking out further Kpop content. Participant 3 responded with, ”It is not really my thing but I can understand the appeal.” What moments stuck out to them often relied on their ability to understand it. A connection was identified between cultural background and points of interest to the participant. Many of the participants tried to relate back concepts of the videos to things related in their cultures. If the content seemed to ”vague” or ”weird” for them they displayed this with more down turned facial expressions, squinting of the eyes, tensed posture, and negative verbal feedback. While as for content that they found relatable was met with a posture that leaned toward the content and an upturned mouth. An example of these two behaviors is shown in the figures below.

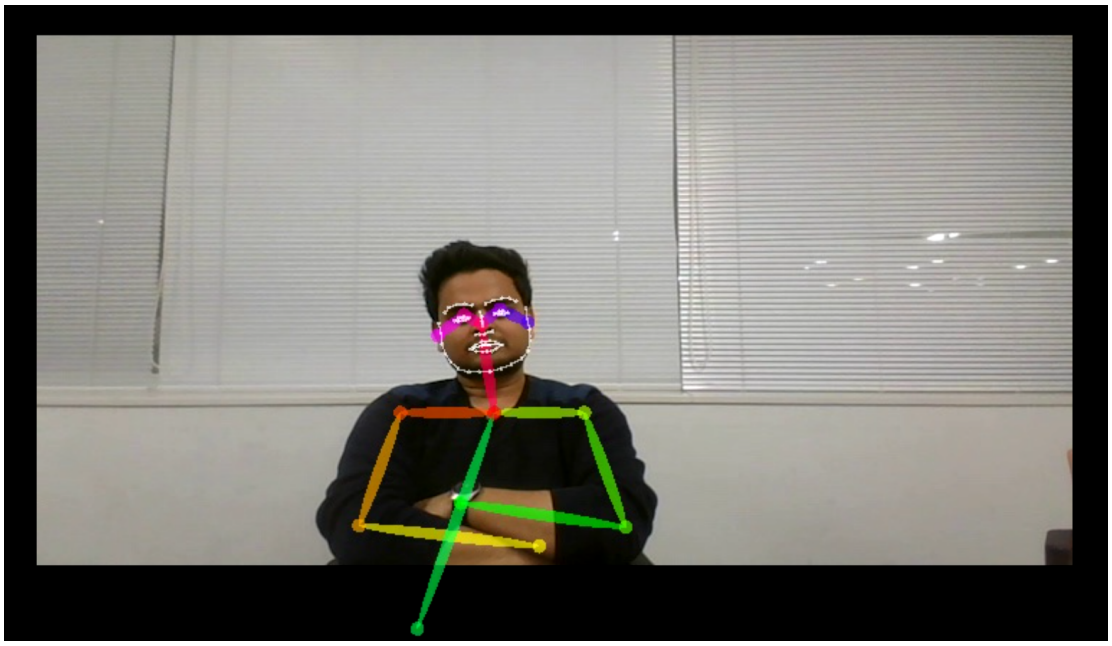


Figure 3.4: P3 in Openpose displaying negative/passive posture.

Figure 3.5: P4 in Openpose displaying positive/active posture.

A correlation was found between religious belief system, sexual education, and

the interpretation of content. Those in more conservative cultures had trouble with more "sexual" or "flirtatious" content. An example of this effect can be shown in these opposing quotes, one from a heterosexual Italian female aged 26 with no religious affiliations, the other from a heterosexual Indian male aged 23 with strong ties to the Muslim religion. Participant 1 when asked about the events in the f(x) video "4 Walls", "All of the members are female. One might be gay? I am not sure. Or two of them are dating. They are all at least very close to each other. It's a story of feeling lost after a break up or tragic event. It was very soft and bittersweet. They where a cute couple. I liked it, yeah." This response can be compared to that of Participant 3, "One of them is a guy right? (after this question being answered with no) But like two of them where in bed together? Are they supposed to be dating? That's weird. I guess it's a love story or like about being confused. I don't know. It was pretty but weird." Both participants liked the video to varying degrees but their understanding and feelings toward the videos content reflected the wide range of human responses do to the influence of cultural. It was evident that those of the heterosexual persuasion preferred videos with the opposite sex in them. The bisexual participant preferred videos of the opposite sex as well. When asked about this they replied, "I think the men look effeminate and therefore they kind of look like both things I like." Females showed a liking to same sex videos replying that, "I can see what I want to be like or like, they are my friends." "I want to be friends with them." "It is easier to relate to them." Male participants did not display a similar response.

Testing showed that the physiological information system to locate key points of interest functioned well. These key points of interest did show correlation to the verbal information responses given to interview questions regarding points of interest. It was possible to see positive physical responses to portions of the content as well as negative or "bored" responses. These positive responses where also reflected in the responses given by participants. Often when a participant named the key points of interest or enjoyment for them it was reflected in the physiological data. The same can be said for the negative physiological responses. When participants stated distaste for a portion of a video it was clearly displayed in the physical data. The hardest state to record though was boredom. As this state of being could also be interpreted as being at rest or simply not feeling an overt

feeling in general. The baseline recording was essential for understanding each persons responses. It was clear while going through the recorded point data that people express themselves in varying degrees that can only be slightly compared to the "average" human response catalog. [36] Making this portion of the results somewhat unreliable. It would require more in depth processing and more precise tools of recording and measure to insure any significant data.

Participants reported difficulty in categorizing the music video's genre both musically and visually. Participants where at first asked to give a genre from the top of their heads. This proved quite difficult for all participants except for one video, BIGBANG's "Fantastic Baby". All participants placed it under the sci-fi fantasy genre and/or post apocalyptic revolution genre. They classified the music style as a club banger. This is similar to the genre it is usually classified under on music distribution sights. It was easier for the other videos when they where given a list of music and movie genres to select from. Even with the list though there was variation in the answers given by participants.

Identifying adjectives, key terms, and the color palette for each video proved easier then picking a genre for participants. Many respond quickly and with multiple responses each time. Adjectives and key terms showed a great level of over lap. Displaying a large number of similar word choices or similes. The color palette showed little diversity. Participants reported similar responses for each video in kind. Answers often reflected a feeling or mood and then participants would describe it by what colors they remembered. For example when Participant 4 was asked about the color palette of Block B's "Her" they replied with, "It had a bright, colorful mood. Like it wanted to be super happy and saturated. Bright vibrant pink, light blue, pastels, and yellow." Another example of this phenomenon is shown with Participant 5's response about the video "I Am the Best" by 2NE1, "This video had a really tough girl, bad girl, shiny vibe. Lots of metal or metallic colors, black, gold and silver, deep purple, and such." A mood followed by the detail description.

3.2.6 Insights

The experiment made clear the gaps in the interview questions. In the next experiment questions about mood would be added as a main inquiry after the

results from the color palette question. Questions about sexuality while interesting deviated from the primary goal. In possible future studies it could prove insightful but for the sake of this thesis it was dropped in the next study. Further study could also be done into gender or race's influence on the interpretation of the videos but for similar reasons to that of sexuality it was dropped for the primary study.

This study lacked an understanding of participants emotional states before and after exposure to the content. Emotional states hold a large effect to participants responses in surveys.(blank) Failing to address this factor left a possibility for bias that could not be afforded. The emotional state could also show how content in the videos affected mood of participants. This information could be added to the mood tag portion of the system giving further depth to its field. Therefore something to measure emotional state was added as a needed dimension in the next study.

It was found that the data presented by Openpose was not as precise as needed for more accurate findings. A need to refine the facial tracking system was deemed necessary if this component was to continue being utilized. Being able to see more minute forms of expression would be necessary for this information to be truly valued. Though the physiological evidence was found useful in providing a correlation between the reported interest participants felt about certain parts of the videos and their physical responses, it no longer remand necessary for future study. Many studies already exist in the realm of physiological response toward content and displaying interest and other motional states. [15,20,22,30] It held no further use or novelty for the study. Rather it presented more a possible diversion and more effort then practical for this thesis's goal. It would not be used for the further gathering of tags for the annotation system.

Further limitations where presented due to the small sample size of the experiment. The universal searching system would require a large arrange of key words, descriptions, and phrases to become tags to feed into the annotation system. Otherwise it would not be able to encompass the large diversity of tags potential users might use to search for content. The experiment did not address the goal of globalization of the tagging system in full. Yes it took into account multiple backgrounds and cultures but it failed to address language barriers. It

would have to be conducted in multiple languages to at least begin addressing this problem.

During the interview it was found difficult to get people to comment on just the visuals without addressing on the qualities of the song. Confirming a certain level of necessity for including the musical components in the examination and creation of the annotation system. In the next experiment steps would have to be taken to examine the musical portion of music videos. The properties of these steps would need to be able to combine smoothly with the data of the visual portion. Lastly the examiner would not be in the room for the next experiment. It was evident that participant's reactions where effected by the presences of another person in the room, creating a biases that needed to be eliminated.

Chapter 4

Tag Identification Study

The goals were formed with the conclusions of the exploratory study in mind and the elements found in the literature review. The Tag Identification Study had four major goals:

- 1: Obtain words, key terms, and phrases users use to describe/search for music video content. Built from the interview questions asked in the first experiment, the new survey would emphasize the participant's use of adjectives, mood, and themes to describe the videos. The questions were directed by the information found in the exploratory research and from the common search engine questions users use to search for music videos online. [25,26,42]

- 2: This experiment would gather data in multiple languages, English, Korean, and Japanese, as a step toward globalization of the tags.

- 3: Responses gathered in this experiment would be added to existing data, to each language's responses, and to the vocal annotations of KAIST. Addressing the duality of music videos by combining data from the three layers of data decided upon, metadata, visual descriptors, and vocal annotation.

- 4: Goal four was to formulate a plan of attack for Piece's development and testing in future works by consolidating the tags generated from the three processes of data collection and then converting the tags into annotations that could be used in it's search system.

4.1. Methods

4.1.1 Procedure

The process for this study was comprised of four parts. After filling out the consent form given participants where sent a link to the Google drive housing the documents/surveys that made up the experiment. This drive also held a form containing the instructions for the participants to follow as they moved through the steps. Step one asked participants to fill out the Positive and Negative Affect Schedule or PANAS-SF for short. This test is made up of two mood scales that measure positive affect and the other measuring negative affect. The scale is used to share relations between positive and negative affect and changes in mood due to external stimuli. Affect in psychology refers to whether you feel emotionally positive or negative in regards to mood. Users are required to respond to 20 items in the test. It uses a five-point scale to measure between the lowest responses, 1 "very slightly" or "not at all" to the highest response, 5 "extremely". Participants answer to what amount they have felt a given feeling within the last week. The span of time can change depending on the use case of the PANAS test such as, right now, today, the past few days, or even a year. The instrument was obtained from the American Psychological Association's website and its agreements for free use in the case of testing and research. Psychologists Watson, Clark, and Tellegen originally created the PANAS scale in 1988. [43] The scale also gives an average positive affect and an average negative affect of human beings as a baseline for study. The test was used twice in the study. Participants were asked to fill out the PANAS test first before moving into the core portion of the study.

Part two consisted of an initial survey collecting the demographic information of each participant. The survey asked for information about participants' music and music video consumption habits. Some examples of questions are, how do they find music/music videos, how often do they listen or watch them, and what kinds of music and music videos do they prefer. It also asked the participants what their experience with Korean pop music and music videos was. Have they listened/watched much Kpop? Do they like to listen/watch Kpop often? Lastly it asked about their ability to understand, speak, and/or read Korean. It was important to understand to what level the participants would be able to under-

stand the content. This does not mean just in the sense of language. The amount participants have been exposed to Kpop culture would affect their comprehension of each video.

In part three participants watched each video one at a time. Music videos for the primary experiment when chosen by the same algorithm used in cross-sectional study one. Slight adjustments where made to the videos selected though. Rather then a pool of ten videos chosen and viewed randomly the videos would be a consistent five. All participants would view the same five videos in the same order. The order is fixed to lower the possibility for variations due to change in order. Due to this limitation of videos, the criteria for selection had to be adjusted. There would be no way to show all facets of the Kpop genre, its contributors, and companies. This time the music video company became less of a critical factor for selection. It was made sure that no company would be represented more then once, but the attempt to distribute evenly between major and minor companies was not of import. Videos continued to be representations of different styles of song and visual style. A balance of genders, groups, and solo artists was preserved but became of less importance compared to the styles used in the video. The focus was narrowed into participants' understanding and ability to describe visual and musical styles shown. When ordering the videos for viewing it was essential that videos with slightly more similar styles where not right after each other. Rather videos with opposing styles where placed back to back in an attempt to help sharpen the contrast. The table below shows the full list of videos chosen for this round of testing.

Table 4.1: Participants

Artist	**Title**
BTS	Fire
Orange Caramel	Catallena
BLACKPINK	Boombayah
f(x)	4 Walls
Jang Deok Cheol	Good Old Days

After each video participants would fill out the survey for that video. They

where asked to answer right afterwards so as to respond when the memory was still fresh. This also kept memories of each video from blending together. A total of five videos would produce a total of five separate surveys. The survey questions where reworded each time to undercut possible misunderstandings in the wording of each question. This helped with curbing repetitive or lazy responses. After analysis of the questions used in the first cross-sectional survey study the surveys where adjusted for more accurate relevant quarries. Questions this time focused more on the moods, aesthetic styles, and the memorable qualities of the content. Leaving behind questions that distract from these topics such as, questions about sexuality and/or gender's effects on perception. Instead the focus became more on language and the descriptors it would bring forth. Questions drove to find key points of interest to the participant. What would they not only notice but, remember. Adjectives, colors, moods, and tones where sought after. The only questions that deviated from this train of thought asked after participants' willingness to view again and/or share each video. Using these two questions to measure true enjoyment of the video and to what degree.

At the end of the study the participants where asked to fill out the PANAS-AF scale for a second time. Reporting how they felt at the end of the experiment after being exposed to the content. The two tests where totaled up and the scores were then compared to each other. Looking to see if there was any positive or negative change to disposition after participating in this study.

4.1.2 Participants

The study concluded with twenty-two English-speaking participants and five Japanese-speaking participants. Due to time constraints the Korean portion of this study was left for future study. In the English-speaking group 77.3 percent where female, 18.2 percent male, and 4.5 percent answered as other. Heterosexual dominated the sexuality category with 86.2 percent of participants, 13.6 percent identified as LGBTQ. Ages ranged from 21 years old to 58 years old, with the highest result 22.7 percent being 22. Nationalities showed a wide diversity, Chinese, Mexican, Italian, Greek, Taiwanese, Thai, Polish, Moroccan, Malaysian, German, Cuban, and American as the highest result at 36.2 percent. All participants of course spoke English. Only 4.5 percent of the participants could only speak En-

glish, all others could speak two or more languages. No participant spoke Korean to any degree. Christian was dominant in the religion category at 45.3 percent. Buddhist came in second at 18.2 percent, none religious third at 13.5 percent, and Islamic in fourth at 4.5 percent. 77.3 percent of participants had lived or are living in another country. Only 22.7 percent have never lived in another country.

4.1.3 Results

Pre-Survey

86.4 percent of participants reported that they enjoy listening to music. The other 13.6 percent stated that they do sometimes. No participants answered no they do not enjoy listening to music. Musical tastes spanned a wide distribution. The top answer from participants was that they enjoyed all or at least most types of musical genres. Of the participants that stated more specific genres, the top responses where Rock, Pop including K, C, and J-pop, Ambient or soundtracks, Classical, Jazz, Indie, and Electronic music genres in that order. Almost all participants find their music online, through music services or corresponding apps. Other methods for finding music mentioned where through friends/family, radio or TV, and social media. The top online services used by participants are Spotify and YouTube, 8tracks making a close third. 54.5 percent of participants often try new things. 22.7 percent try new things all the time and another 22.7 percent do sometimes. When asked if it was important to them that the language of a song was one they could understand 13.6 percent said yes, 54.5 percent do not mind at all, and 31.8 percent said sometimes it matters. This feeling was mainly prevalent in how the participants choose to consume musical content. For some they like it more for the beat or for back ground music, "It doesn't matter if I can understand the language, as long as I enjoy the beat or rhythm." For others the lyrics help with emotional connection to the work. "If I understand the meaning behind it, I get more attached to that song cause sometimes I can relate it to my current mood or situation I am in." As this participant demonstrates these two desires are not exclusive to the individual. "Most of the time music is like background music for me so I don't always listen closely to the lyrics. But when a song strikes me with lyrics eg that fit my current situation or touches me personally

then I like listening to a song because of the lyrics. Also I'm used to listen to not only English music that I was able to understand quite early in my life but also Japanese/Korean/Thai music ever since I was a kid when I obviously didn't understand any of those languages." Mood or intention while listening plays a large part in the need for understanding on a language level. Sometimes beat or sound is enough to satisfy the conditions for connection.

63.6 percent of participants watch music videos sometimes. 22.7 percent do not watch them at all and 13.6 percent watch them frequently. This came as a surprise. The amount of participants who stated that they do not watch music videos whatsoever in the survey were higher then what was predicted in the hypothesis. Their answers as to why they do or do not like music videos revealed the main hurtle for those who do not like to watch videos. "Videos require more focus, I usually listen to music while doing something else. Sometimes I prefer to concentrate on just the music, too." "I rarely find them that interesting. Short attention span." "No time to sit down and watch music videos, usually listen to music while walking, being on the train, or doing something else that needs my attention (work, study etc.), don't want to use Internet data for videos in general while on the go." These quotes point to the main reason for not consuming music video content, they felt that they had no time or desire to spend their time on watching music videos. Music video, unlike music alone, requires the person to go out of their way and truly focus on the product. In the participants opinions, music can be enjoyed while multitasking. For those that do watch music videos they find enjoyment in the themes, aesthetics, and visual impact that videos can have. 43.6 percent of the participants like videos with choreography, the more elaborate the better. The main response was for the ability of music videos to convey a story. "Sometimes they convey important messages and are very creative." "I like when they tell a story that's not evident from the lyrics alone, and I like when there is interesting dance choreography." "Gives a new perspective on the music mainly. But sometimes it's so visually effective that it helps me understand the message of the artist more." As made evident from some of these quotes from the participants music videos can help provide explanation about the song or the artists vision, provide more context, and/or present a new interpretation of the song then what they had. For those who enjoy music videos, the content is mainly

found through searches on YouTube, social media, and/or friends. Often sharing the content they enjoyed within these circles.

The last three questions in the pre survey asked after the participants level of exposure, enjoyment, and understanding of Korean pop music. Of participants 36.4 percent have never listened to K-pop before, 31. 8 percent have rarely, 9.1 percent have sometimes, 18.2 often listen to K-pop, and 4.5 percent listen to it all the time. For those that listen to K-pop music most began listening to it around the years of 2015-2016. Some went back as far as beginning in 2001, 2006, and 2008. Making most listeners quite recent consumers. When asked about their consumption of music videos 4.5 of participants watch them all the time, 9.1 percent watch them often, 40.9 percent do sometimes, 13.6 percent rarely, and 31.8 percent do not watch them at all. Again none of the participants can speak, read, or understand Korean beyond at most a few scattered words. This concluded the pre-survey portion of the experiment.

PANAS Test

The PANAS test results where calculated individually by adding up the totals from the ten positive affect questions and the ten negative affect questions then comparing these to the mean score of those that take this test. Both tests where scored then the four scores where compared to each other looking for any changes to the numbers. The differences in scores was recorded and then a designation of positive change or negative change was recorded. If the positive score went up this was marked as a positive change. If the negative score went up this was marked as a negative change. If the positive change out weighed the negative change it was determined a positive over all change in affect for the participant. The table below holds for examples of this record.4.2

Table 4.2: PANAS Scores Examples

Name	**PANAS Scores 1**	**PANAS Scores 2**	**Differences**	**Positive Change**
P2	22P 33N	26P 18N	+4P -15N	Yes
P11	40P 28N	31P 19N	-1P +3N	No
P10	33P 20N	39P 19N	+6P -1N	Yes
P18	32P 16N	31P 19N	-1P +3N	No

The scores of more positive or negative affect from the individual tests is not what was important in the measurement. These can be seen as more like baselines of the participants. The change is what is important. A difference of +,- or +,+ equals a positive affect change overall. A difference of -,- or -,+ equals a negative affect change overall. Of twenty two participants the split was almost equal with 12 positive affect changes and 9 negative affect changes. One participant had no change in affect whatsoever. Significance was determined low. With the totals so close it is necessary to further test the affect of exposure to the content to determine any significant change.

Post-Survey

Each participants responses we're separated by video. Using the Natural Language Toolkit developed by Bird et al. [5], sentences from each participant's responses where broken down into their grammatical parts. Specifically looking for adjectives, nouns, and phrases that could be utilized for creating tags. In the related work section it was discussed that adjectives and nouns tend to be the primary words to use when searching for content so that became a primary focus for information gathering. [25] After answers where broken down into these component parts, the words we're compared to find commonly used terms and descriptors. All repetitions where eliminated by merging together any similar or redundant words to create a segment-level annotation in the system. A hierarchal cluster was created as a result of similarity between tags. After collecting segment-level tag annotation data finally, we utilized the auto tagging results for content-based music search and recommendation.

4.1.4 Insights

The study reinforced a need to focus on the human element of developing tags. Piece has to be social, flexible, malleable evolving with the people. Tags rely on language. Therefore Piece's tags have to be capable of shifting along side languages. By allowing the users to generate their own tags, organize as they please, and share their creations Piece might have the ability to keep up with the ever evolving beast that is language. Mapping similes between languages, such as

house equals in Japanese, builds a capability for cross-cultural search-ability. So if a user does not speak the language of the video they can search in their native language for the content. During interviews

Chapter 5

Proof of Concept Test Study

Now that tag identification had been refined during the tag identification study it was time to look into the practical application of using natural language as a form of generating tags.

- 1: Obtain words, key terms, and phrases users use to describe/search for music video content. Guided by the identification of which descriptors and in what ways people from the tag identification study describe content.
- 2:Focus primary on participants descriptions.
- 3: Compare answers given in one interview with those given in a secondary interview given after a designated period of time. Looking into descriptors or moments that more more retained.
- 4: Compare participants key terms to see what, if any, correlated or matched. Using this comparison to build a rudimentary testable tagging hierarchy system.
- 5: Analyze given responses for application in creating user search interface.

5.1. Methods

5.1.1 Procedure

Participants began by giving their age, nationality, and gender. This study did not focus on demographic info and therefore it was kept to a minimum. The study was separated into two days, about 20 min each day. Day one saw that participants watched two Kpop videos, nether of the videos where used in previous studies. After each video the participant was asked about what they could remember of the video. They where first prompted with only this question. Once answered more specific aspects of the videos where asked to be recalled by the participants. This included what they could remember of the color palette, the mood or feeling of the piece, adjectives they would use to describe the video, and the genre they would give to the video and the music. The questions sought for the participant to describe the video as much as they could both by what first came to mind and by prompting. On day two participants where asked the same questions as the day before but without watching the videos again. Testing what they could recall of the videos. Their ability to remember though was not the primary objective. Rather it was to compare their answers to see if the word choices, descriptions, and what they remembered on their own and with prompting reflected their answers from the day before. If their answers where similar and shared many of the same markers it would serve as a good sign that natural language could be utilized as a means of tag generation for visual media. If there where no or little similarities it would help to show a filure int he system. The answers given by each participant was compared, interviews one and two. After this comparison the two anwers where broken down to its key terms. These key terms where then compared against each other, participant to participant, to see if there was overlap in the participants answers. Overlapping terms where then flagged to be top in the tag heirarchy. Terms that overlapped with a few other answers where placed in the secondary catigory. Finally key terms that did not overlap what-so-ever where placed in the tershiary catigory.

5.1.2 Participants

Total participants in this study totaled twenty three in all. The cross-section represented eight countries. Participants ages ranged from 18-56 years old. Two participants where very familiar with Korean pop music, the rest had limited to no knowledge of Kpop. Fourteen participants identified as female, six identified as male, three identified as other. All participants spoke English fluently.

5.1.3 Results

First the first and second interviews where compared to each other. Below displays examples of the recorded answers. Two participant's answers for the first video and two participant's answers for the second video. A note on the language of the answers. Participants where asked to answer as free flow as they wanted. If an idea or thing popped into their head they where encouraged to just say it, let it flow. Aiding in more natural language responses. Interviews where recored by the researcher and then transcribed. Leaving out more conversational or repettitve responses.

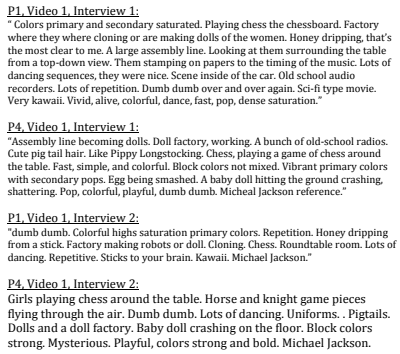


Figure 5.1: Comparison of 2 participants answers in regards to the first video.

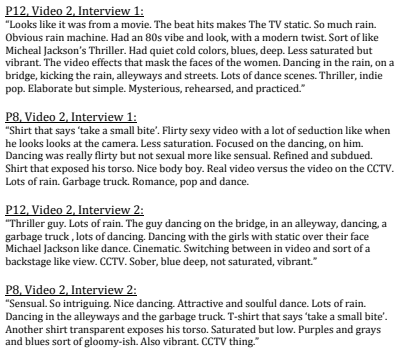
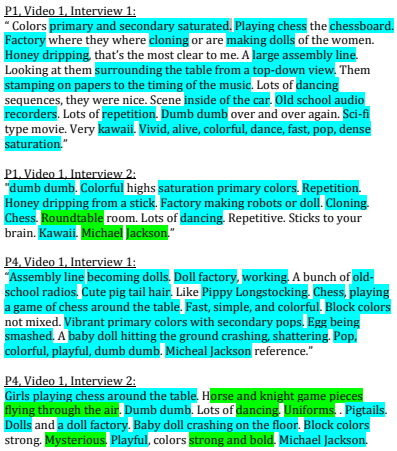


Figure 5.2: Comparison of 2 participants answers in regards to the second video.

A trend emerged immediately that responses to the second interview where more concise and contained far less details. Terms that where mentioned in the first interview where highlighted if seen as possible key terms but if they did not arise in the second interview they where placed into the secondary category of the tagging hierarchy immediately. They would be moved down to the tertiary category if they failed to match anyone else's responses. Responses given in the secondary interview where placed at higher value. Given that they where what people could remember after a period of time. It would be less likely that users of Piece would be searching for content they have just seen or are very fresh in the mind. Often it would be more necessary to search with what bits can be remembered over a period of time. Therefore the most natural of the language would be

the language used after this period of rest. Below is an example of these responses being compared. Blue highlights indicate key terms/phrases used in both the first and second responses. Green highlights are possible key terms/phrases that where not given in the first interview.



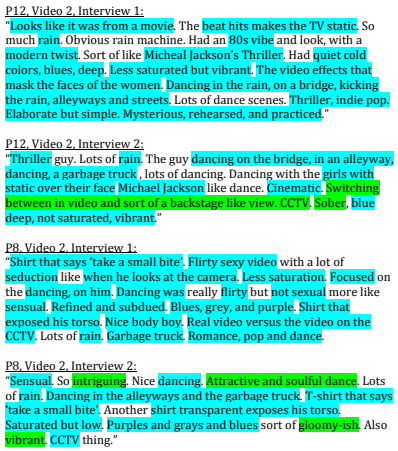
P1, Video 1, Interview 1:
" Colors primary and secondary saturated. Playing chess the chessboard. Factory where they where cloning or are making dolls of the women. Honey dripping, that's the most clear to me. A large assembly line. Looking at them surrounding the table from a top-down view. Them stamping on papers to the timing of the music. Lots of dancing sequences, they were nice. Scene inside of the car. Old school audio recorders. Lots of repetition. Dumb dumb over and over again. Sci-fi type movie. Very kawaii. Vivid, alive, colorful, dance, fast, pop, dense saturation."

P1, Video 1, Interview 2:
"dumb dumb. Colorful highs saturation primary colors. Repetition. Honey dripping from a stick. Factory making robots or doll. Cloning. Chess. Roundtable room. Lots of dancing. Repetitive. Sticks to your brain. Kawaii. Michael Jackson."

P4, Video 1, Interview 1:
"Assembly line becoming dolls. Doll factory, working. A bunch of old-school radios. Cute pig tail hair. Like Pippy Longstocking. Chess, playing a game of chess around the table. Fast, simple, and colorful. Block colors not mixed. Vibrant primary colors with secondary pops. Egg being smashed. A baby doll hitting the ground crashing, shattering. Pop, colorful, playful, dumb dumb. Micheal Jackson reference."

P4, Video 1, Interview 2:
Girls playing chess around the table. Horse and knight game pieces flying through the air. Dumb dumb. Lots of dancing. Uniforms . . Pigtails. Dolls and a doll factory. Baby doll crashing on the floor. Block colors strong. Mysterious. Playful, colors strong and bold. Michael Jackson.

Figure 5.3: Example of identification of key terms in regards to the first video. 2 Participants



P12, Video 2, Interview 1:
"Looks like it was from a movie. The beat hits makes the TV static. So much rain. Obvious rain machine. Had an 80s vibe and look, with a modern twist. Sort of like Micheal Jackson's Thriller. Had quiet cold colors, blues, deep. Less saturated but vibrant. The video effects that mask the faces of the women. Dancing in the rain, on a bridge, kicking the rain, alleyways and streets. Lots of dance scenes. Thriller, indie pop. Elaborate but simple. Mysterious, rehearsed, and practiced."

P12, Video 2, Interview 2:
"Thriller guy. Lots of rain. The guy dancing on the bridge, in an alleyway, dancing, a garbage truck , lots of dancing. Dancing with the girls with static over their face Michael Jackson like dance. Cinematic. Switching between in video and sort of a backstage like view. CCTV. Sober, blue deep, not saturated, vibrant."

P8, Video 2, Interview 1:
"Shirt that says 'take a small bite'. Flirty sexy video with a lot of seduction like when he looks at the camera. Less saturation. Focused on the dancing, on him. Dancing was really flirty but not sexual more like sensual. Refined and subdued. Blues, grey, and purple. Shirt that exposed his torso. Nice body boy. Real video versus the video on the CCTV. Lots of rain. Garbage truck. Romance, pop and dance.

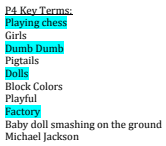
P8, Video 2, Interview 2:
"Sensual. So intriguing. Nice dancing. Attractive and soulful dance. Lots of rain. Dancing in the alleyways and the garbage truck. T-shirt that says 'take a small bite'. Another shirt transparent exposes his torso. Saturated but low. Purples and grays and blues sort of gloomy-ish. Also vibrant. CCTV thing."

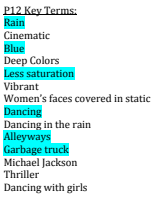
Figure 5.4: Example of identification of key terms in regards to the second video. 2 Participants

Terms highlighted in green where answers given in the second interview but not in the first. These terms proved an obstacle. They where remembered at a later time but should they hold the same weight as those given on both days? We took the approach that like those given in the first interview but not in the second they where not good candidate for the primary of the tag hierarchy. Rather they would better suit the secondary layer until compared with other participants

answers. Continuing with the same examples in the next figure you will see the key terms that I've been broken down from both answers from each interview. They have now been compared to other participants answers to see if there's any correlation between their language choices.

Figure 5.5: Matching terms from two participants. Video 1.





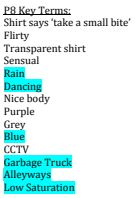


Figure 5.6: Matching terms from two participants. Video 2.

Once compared the terms that arose from several participants answers where kept in the primary level of the tag hierarchy. Those terms that arose and matched with only a few other answers from participants where moved down to the secondary level. Those that match no other answers where moved down to the tertiary level. It was found that some of the answers given in the secondary interview did match with other answers and where kept in the secondary level. Those of the ones that did not were moved down to the tertiary. Through this process

we were able to find twelve key terms for the first video and eight for the second video. The secondary level was much larger at sixteen key terms for the second video and twenty four for the first video. The tertiary level, being that it had lesser constraints, totaled higher but only by a slight margin twenty five for the first video and twenty one for the second video.

5.1.4 Insights

We found that when participants where given prompts they had an easier time recalling information. When continuing the development of Piece it may be wise to take this into account. Giving users the ability to search by stream of consciousness, by detailed account of what it is they are looking for, as well as advanced search options for looking into these particular categories. Possibly with drop down options or checking boxes for the ones that the user is looking for. These categories should include but are not limited to gender of the artists, mood or feeling, Color palette, genre of music and movies, styles, and more specific metadata. Natural language showed that it is a reliable source for tag identification. Auto generation could be created from identifying the descriptors used in the natural language of users to help further facilitate this form of tagging. Continuing this study, as done with the tag identification research, in multiple languages is paramount to furthering our understanding of natural language as a source system for tagging. Even more so when we look to combine languages so that users may search in their native tongue while still pulling up results that might be with tags in another language.

Chapter 6

Combining with KAIST's work

Simultaneously while this process was being run, our colleagues in KAIST (the Korean Advanced Institute of Science and Technology) where working on the vocal analytics of the songs from the videos in study one and study two. Professor Nam and his masters student Chae Lin Park engineered a program called Music Galaxy Hitchhiker, that allows users to navigate through albums in a galaxy like display. One can see the elements that connect constellations of albums and songs. The connections are based off of their analyses of five main features of the vocal qualities of a song. It gives data on the waveform, beat tempo, RMS energy, spectrogram, onset strength, energy, and tracking the beat. Waveform refers to the shape projected on a graph showing the varying quantities over time of a shape or form. It can refer to a signal, a wave moving in a physical medium, or an abstract representation in general. Beat is a regular repeating pulse that is set underneath the musical pattern. Often referred to as the "heartbeat" of a song. Tempo is the speed with which a musical pattern is played back, measured in beats per minute (BPM). Beat tempo is the speed at which the beat of a musical piece is played back. Beat tracking records the tempo and the wavelength of the beat throughout the song. Root mean squared or RMS is the calculation of the square root of the mean square. The calculation of this in music measures the values of the voltage and current in waveforms. It takes into account the resistive load caused by the waves power or energy. This calculation affects the impact of when a wave is passed through an amplifier or speakers. Spectrograms are visual representations of the spectrum of frequencies of sound on other signals as they

vary with time. They may also be referred to as sonographs, voice points, or voicegrams. The spectrograms generated by Professor Nam and Chae Lin Park display the onset strength and the energy of a song as well as displaying the other information mentioned above.

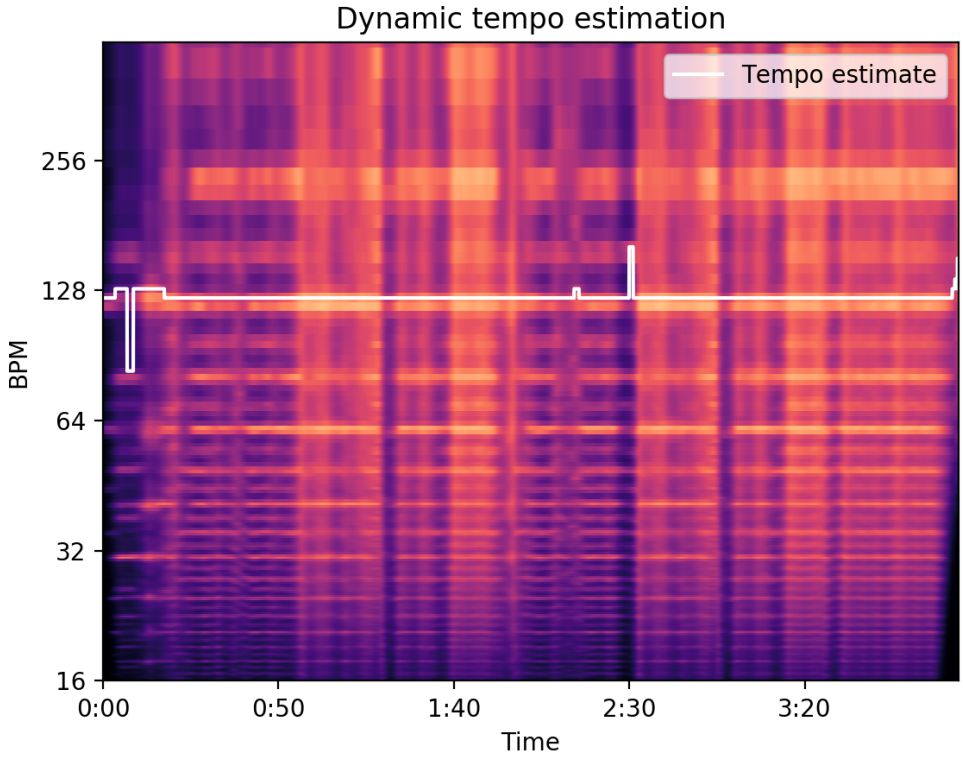


Figure 6.1: Example of Tempo recorded output from KAIST Analysis

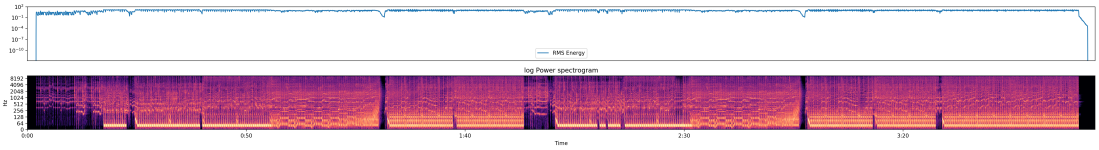


Figure 6.2: Example of RMS recorded output from KAIST Analysis

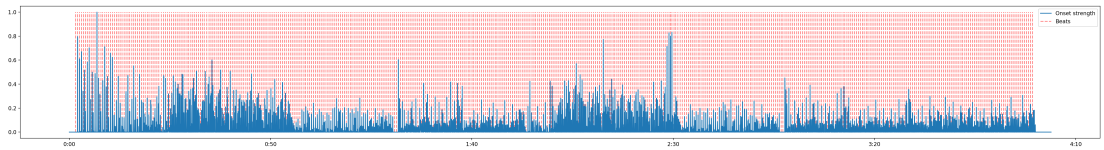


Figure 6.3: Example of recorded beat tracking from KAIST Analysis

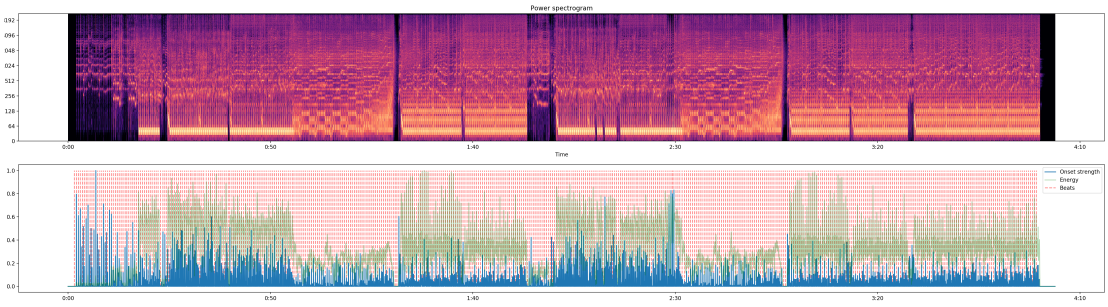


Figure 6.4: Example of Spectrogram recordings of energy output from KAIST Analysis

There work took the recorded waves and placed them to labels they had developed previously for their initial study. [17] These labels consisted of seventy descriptions about vocal qualities. This included the song as well as any script portions in the video. Previously they had worked on an initial research project in creating a vocal annotation tagging system. This project took the output displayed in terms of spectrograms and marked the elements into terms used in describing vocal qualities. Then incorporated into Chae Lin Park's Music Galaxy Hitchhiker. Their methods where then employed on the videos selected in our experiments. Currently they are breaking down the spectrograms into the terms. These terms will be combined with this thesis's findings to add another dimension to the segment-level tag annotation data system. This included the song as well as any script portions in the video. Currently they are breaking down the spectrograms into the terms. These terms will be combined with this thesis's findings to add another dimension to the segment-level tag annotation data system. The final goal of this collaboration is to build an arm of Chae Lin Park and Professor Nam's Music Galaxy Hitchhiker program for music video navigation.

Chapter 7

Conclusion

Piece is above all a thesis on providing a piece of the puzzle when it comes to how to search and tag visual media. It is not only plausible but an effective tool for future generations as a recourse to gather and share their visual taste. It is a viable tool for companies to get their music video works out to the consumer in a way more tailored for crossing cultures and for the user to be able to utilize their natural language as a means for access. The final research validates the tagging system while providing a rudimentary hierarchy that can be seen as guidelines for further building on this form of navigation. With combining the Music Galaxy Hitchhiker system of professor Nam and his students at KAIST Piece and its theoretical process for obtaining and sharing music video media in the K-pop genre has become a closer reality. By using people's natural modes of speech, from multiple countries, we can build a user friendly system that is not hindered by language. Now when you can't find the name of the song, the artist, the company you will not be stuck adrift.

Music videos are a marriage of visual and musical content. Despite this harmonious combination music videos are left to be categorized by their musical qualities alone. Those seeking music videos work with a limited tool set to find their chosen content threw methods that completely ignore the visual aspects of the media. The aim of this thesis was to contribute to the initial development of an augmentation to pre-existing music recommendation systems. To extend the capabilities of systems by making use of how people speak, in the language they use. No matter what language you use it will be possible to find what you want to

look for. Extending the systems range with the addition of a music video search ability by integrating tags inspired by the visual elements of videos. Giving the users the ability to create their own tags and uses. Presenting Piece, a tool built on the natural language of the people to create a natural system. The intellectual property of Piece endows the system the ability to search for music videos by their visual qualities, metadata, and by user suggested tags. No longer will you have to scramble in the dark. As long as you have a piece of the puzzle, Piece will give you a peaceful search experience.

7.1. Limitations

We where consistently limited by our participant numbers. Finding a way to reach more users for testing would be in great demand if Piece and natural language tagging where to ever be moved into a workable system. Sadly without a form of compensation and low access to both the Japanese and Korean populations it was not possible at this time. Language barrier also played a factor in limiting findings. Since the lead researcher was limited to just an understanding of English and limited Japanese and Korean all tests in the other tests had to be farmed out to volunteer researchers. This included translation of answers back into English for comparison. Since the researchers where not professional translators all translations had to be handled with care to not over emphasize there possibility as corresponding matches. The over all study would benefit from conducting a more long term, large cross-sectional version of the last two studies.

7.2. Future Works

As for future applications and research we hope to move onto, as with our colleagues at KAIST, developing an AI (artificial intelligence) that is able to annotate and analyze music videos based on the knowledge fed to it from its users and begin automatically tagging videos as theyre uploaded into the Piece system. This will mean furthering a long term study like those of the tag identification study and the key term identification study. They will require conduction in multiple languages with far larger sample sizes to accurately represent the large

population of music video search engine users. A strong push toward a growing understanding of further languages that could be applied to the user generated tags and then mapped to corresponding tags in other languages is a major push for Piece's future. Adaptation of the key term/tag identification methods described in chapter six into an automatic process is a clear necessity to make Piece workable in the real world. It could also be made possible for Piece to have its own portal or RSS system in the future.

Acknowledgements

This one is to my mother, the only woman who can keep me sane and others alive. My editor, cheerleader, best critic, and friend. To my dad, the one who forged me on the comedy fires of the Marx Brothers, Monty Python, and Pink Panther. You taught me to be strong yet sensitive. You showed me how to keep others laughing and smiling. The two of them together raised a crafty crafty daughter. To Brittney, you did so much to see me and to support me. To my friends back home. To my partner in crime Leigh, who took care of our sweet baby boy while I sought education. To all my friends here on Japan's soil who toiled and worked as we lifted each other up. To Ploy, Stephanie, and Tanner who listened and listened again. To Itsuki and Miho who helped be to make a home here in Japan. Thank you to Akira and Yuni who reminded me what I love about myself, gave me permission to. Thank you to Professor Kai Kunze, mainly for the food. Thank you to Matthew sensei who called to help me even in one of the hardest moments of his life. You kindness saved me. Thank you to my upperclassmen George and Ben, I could not have done it without you. Finally to Terry Pratchett who taught me how to see humor in this world. Who showed me how to make glass shine like diamonds. To all of you thank you.

References

[1] Aucouturier, J.-J., and Pachet, F. Representing musical genre: A state of the art. *Journal of new music research 32*, 1 (2003), 83–93.

[2] AUTODESK. Faq: What is the difference between a tag and an annotation?, Oct. 31 2017.

[3] Bainbridge, D., Cunningham, S. J., and Downie, J. S. How people describe their music information needs: A grounded theory analysis of music queries.

[4] Bennett, A., et al. *Popular music and youth culture: music, identity and place.* Macmillan Press Ltd., 2000.

[5] Bird, S., Klein, E., and Loper, E. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[6] Chen, H.-C., and Chen, A. L. A music recommendation system based on music data grouping and user interests. In *Proceedings of the tenth international conference on Information and knowledge management*, ACM (2001), 231–238.

[7] Cunningham, S. J., Reeves, N., and Britland, M. An ethnographic study of music information seeking: implications for the design of a music digital library. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries*, IEEE Computer Society (2003), 5–16.

[8] Downie, J. S. Music information retrieval. *Annual review of information science and technology 37*, 1 (2003), 295–340.

[9] Fabbri, F. Browsing music spaces: Categories and the musical mind.

[10] Foundation, K. *Global Hallyu 2015: Charting the Popularity of Korean Culture in 110 Countries, Vol.1-4*. Korean Foundation, 2015.

[11] Foundation, K. *Global Hallyu 2017: Charting the Popularity of Korean Culture in 112 Countries, Vol.1-4*. Korean Foundation, 2017.

[12] Fu, H., and Fan, Y. Music information seeking via social q&a: An analysis of questions in music stackexchange community. In *Digital Libraries (JCDL), 2016 IEEE/ACM Joint Conference on*, IEEE (2016), 139–142.

[13] Futrelle, J., and Downie, J. S. Interdisciplinary communities and research issues in music information retrieval. In *ISMIR*, vol. 2 (2002), 215–221.

[14] Gill, P., Arlitt, M., Li, Z., and Mahanti, A. Youtube traffic characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, ACM (2007), 15–28.

[15] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[16] Incorperated, T. S. Tobii eye tracking systems.

[17] Kim, K. L., Kum, S., Park, C. L., Lee, J., Park, J., and Nam, J. Building k-pop singing voice tag dataset: A progress report. In *The 18th International Society for Musical Information Retrieval Conference*, ISMIR (2017).

[18] Kim, Y. *The Korean wave: Korean media go global*. Routledge, 2013.

[19] Kreitz, G., and Niemela, F. Spotify–large scale, low latency, p2p music-on-demand streaming. In *Peer-to-Peer Computing (P2P), 2010 IEEE Tenth International Conference on*, IEEE (2010), 1–10.

[20] LaFrance, M., and Broadbent, M. Group rapport: Posture sharing as a nonverbal indicator. *Group & Organization Studies 1*, 3 (1976), 328–333.

[21] Lavranos, C., Kostagiolas, P., Korfiatis, N., and Papadatos, J. Information seeking for musical creativity: A systematic literature review. *Journal of the Association for Information Science and Technology 67*, 9 (2016), 2105–2117.

[22] Lee, H. C., Hong, T. T., Williams, W. H., Fettiplace, M. R., and Lee, M. J. Method and system for measuring and ranking an engagement response to audiovisual or interactive media, products, or activities using physiological signals, July 1 2014. US Patent 8,764,652.

[23] Lee, J. H., Cho, H., and Kim, Y.-S. Users' music information needs and behaviors: Design implications for music information retrieval systems. *Journal of the Association for Information Science and Technology 67*, 6 (2016), 1301–1330.

[24] Lee, J. H., Choi, K., Hu, X., and Downie, J. K-pop genres: A cross-cultural exploration. In *Proceedings of the 14th Conference of the International Society for Music Information Retrieval (ISMIR)*, The International Society for Music Information Retrieval (ISMIR). (2013).

[25] Lee, J. H., and Downie, J. S. Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In *ISMIR*, vol. 2004, Citeseer (2004), 5th.

[26] Lee, J. H., Downie, J. S., and Cunningham, S. J. Challenges in cross-cultural/multilingual music information seeking.

[27] Logan, B. Music recommendation from song sets. In *ISMIR*, Citeseer (2004), 425–428.

[28] McEnnis, D., and Cunningham, S. J. Sociology and music recommendation systems. In *ISMIR* (2007), 185–186.

[29] McKinney, M., and Breebaart, J. Features for audio and music classification.

[30] Mota, S., and Picard, R. W. Automated posture analysis for detecting learner's interest level. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, vol. 5, IEEE (2003), 49–49.

[31] Norowi, N. M., Doraisamy, S., and Wirza, R. Factors affecting automatic genre classification: an investigation incorporating non-western musical forms. In *Proceedings of the International Conference on Music Information Retrieval*, Citeseer (2005), 13–20.

[32] Openpose.

[33] Pachet, F. Content management for electronic music distribution. *Communications of the ACM 46*, 4 (2003), 71–75.

[34] Pampalk, E., and Gasser, M. An implementation of a simple playlist generator based on audio similarity measures and user feedback. In *ISMIR* (2006), 389–390.

[35] Petrelli, D., Beaulieu, M., Sanderson, M., and Hansen, P. User requirement elicitation for cross-language information retrieval. *The New Review of Information Behaviour Research–Studies of Information Seeking in Context 3* (2002), 17–35.

[36] Shan, C., Gong, S., and McOwan, P. W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing 27*, 6 (2009), 803–816.

[37] staff writer, K. Youtube kpop music videos views january 2018, Jan. 1 2018.

[38] Tardón, L. J., Sammartino, S., Barbancho, I., and Barbancho, A. M. A multidimensional environment for the exploration of musical content. *Computer Music Journal 36*, 3 (2012), 73–83.

[39] Taylor, P. G. Press pause: Critically contextualizing music video in visual culture and art education. *Studies in Art Education 48*, 3 (2007), 230–246.

[40] Tzanetakis, G., and Cook, P. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing 10*, 5 (2002), 293–302.

[41] Vaessens, B. *Expression of music preferences: how do people describe popular music that the want to hear*. PhD thesis, Master thesis, 2002.

[42] Vignoli, F. Digital music interaction concepts: A user study. In *ISMIR*, Citeseer (2004).

[43] Watson, D., Clark, L. A., and Tellegen, A. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology 54*, 6 (1988), 1063.

Appendix

A. Tag Identification Surveys

The following are copies of the surveys participants where asked to fill out while participating in the tag identification experiment.

THE OHIO STATE UNIVERSITY

Positive and Negative Affect Schedule (PANAS-SF)

Indicate the extent you have felt this way over the past week.		Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
PANAS 1	Interested	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 2	Distressed	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 3	Excited	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 4	Upset	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 5	Strong	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 6	Guilty	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 7	Scared	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 8	Hostile	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 9	Enthusiastic	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 10	Proud	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 11	Irritable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 12	Alert	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 13	Ashamed	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 14	Inspired	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 15	Nervous	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 16	Determined	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 17	Attentive	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 18	Jittery	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 19	Active	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PANAS 20	Afraid	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Figure 7.1: PANAS test used in the Tag Identification Experiment, pulled from the resources of the American Psychological Association.

General:

Nationality:

Age:

Gender:

Male	Female	Other

Sexuality:

Language(s) you speak:

Religion:

Have you lived in another country? If yes, where?

Part 1:

Do you enjoy listening to music?

Yes	No	Sometimes

If yes, what kinds of music?

If yes, where do you find your music? (Online, Friends, Etc.)

Figure 7.2: Pre-survey used in the tag identification experiment.

Post Survey

Thank you for participating in our experiment. Please fill out each question to the best of your ability. There are no right or wrong answers. This is based on your opinions and memory. The more detailed you can be the better. Thank you!

Video 1 Fire by BTS:

1. What was the basic plot of the video?

2. What was the most significant part of the video to you?

3. Did anything in particular stand out to you?

4. What was the theme of the video?

5. What was the color palette of the video?

6. What was the mood of this video?

7. Please give some adjectives to describe this video.

Figure 7.3: Post-survey used in the tag identification experiment.