A Comparison of musical elements in J-pop and VOCALOID music

Bowen Dong (a), Arthur Claire (b)

(a) Georgia Institute of Technology, United States, bdong46@gatech.edu, (b) Georgia Institute of Technology, United States, claire.arthur@gatech.edu

This study aims to investigate the differences in melodic characteristics between VOCALOID music and J-POP music, focusing on pitch, intervals, and singing speed. We manually collected and created a corpus of J-POP and VOCALOID songs and conducted a statistical analysis using humdrumR. The results of our data analysis revealed that VOCALOID singers tend to have a slightly higher average pitch than J-POP singers, and generally sing at a faster pace compared to J-POP singers, but do not use larger intervals in their main melodies. We hope this research can fill the gap in the study of the emerging online subculture community of VOCALOID, providing valuable insights into the unique aspects of this musical genre.

Keywords: VOCALOID music, J-pop music, Computational musicology, Corpus research, Synthesis voice, ICMPC17

1. Introduction

VOCALOID is a voice synthesis software for electronic music production. VOCALOID music emerged in 2007 along with the appearance of the most famous library of VOCALOID, Hatsune Miku (Donna Li, 2021). Currently, VOCALOID music refers not only to songs that use the sound of VOCALOID software as lead vocals but also to any popular song that makes use of a synthetic voice as lead vocals, including those made with other softwares such as UTAU, Synthesizer V, and Cevio as lead vocals. They are collectively known as VOCALOID music simply because VOCALOID was one of the first to appear and is the most representative. Because the most representative VOCALOID libraries initially supported only Japanese, the aesthetic of this community is heavily influenced by Japanese pop music, though not identical.

Since its emergence in 2007, an online subculture community has developed. The influence of VOCALOID music (henceforth "VOCALOID") has grown significantly has grown significantly over the past fifteen years. For example, the most popular VOCALOID track on Youtube, "*Goodbye Declaration*," has already received over 100 million views.

Unlike J-pop music, amateur musicians are the primary distributors of their works in the VOCALOID music community. This community serves as a springboard for amateur musicians to launch their careers in the professional music industry. Over the past decade, many examples of amateur musicians have used this springboard to jump into the professional music production world. One of the most famous examples is Genji Yonezu, who uploaded original songs using VOCALOID between 2009 and 2012 and subsequently debuted as a singer-songwriter after gaining notoriety. His masterpiece, "*Lemon*," has received over 100 million plays on Spotify and has been ranked #1 on Oricon's publicly available digital singles chart for six weeks (Billboard Japan, 2019).

1.1.J-Pop Music

J-pop emerged as a term to describe Japanese popular music in 1988 via the founding of a Japanese FM radio station (Finan, 2020). Nowadays, J-pop music refers to all Japanese popular music other than Enka. (Enka is a unique Japanese song that combines the singing styles of Japanese folk artists from the Edo period and incorporates the national moods of various parts of Japan. It is the transition between Japanese classical art and modern pop music.)

1.2. Motivations

This research project aims to contribute to a better understanding of the preferences of the VOCALOID community and the J-POP music community, given the significant role that VOCALOID music plays in the Japanese music industry today. Moreover, J-pop has had a large influence on other pop music traditions, particularly in other parts of East and Southeast Asia (Finan, 2020; Bourdaghs, 2012). Hence, studying the characteristics of VOCALOID compositions as a subculture under J-pop is an intriguing area for research.

Another important motivation for this project is the opportunity to create a representative dataset that can be utilized for musicological research on VOCALOID. As the VOCALOID community's influence continues to grow, there is an increasing interest in researching this genre. As recently as December 2022, Billboard Japan announced the launch of a ranking dedicated to VOCALOID music. However, reliable datasets for VOCALOID music research are currently unavailable. Therefore, this project aims to fill this gap by creating a high-quality dataset that researchers can use to study the different elements of VOCALOID music, such as pitch, melody, rhythm, and singing speed, or other musicological features.

Through informal observation, it seemed that

VOCALOID songs partly create their own unique identity separate from J-pop songs by the use of features and characteristics that appear to differentiate it from those of J-pop, such as an extremely fast singing speed. Specifically, the use of a synthetic voice affords certain musical possibilities that may be difficult (if not impossible) to carry out with the human voice. We conjectured that if an intentional component of creating a VOCALOID music "sound" was through the use of unnatural melodic features, that we would be able to observe a difference in those features via corpus analysis. Specifically, based on this conjecture, we hypothesize that overall, VOCALOID singers:

1. Sing higher than J-POP singers.
2. Use more frequent and larger melodic intervals than J-POP singers.
3. Sing faster and have more frequent note onsets (per unit time) than J-POP singers.

1.3. Related works

In 2012, Sasaki Asuka et al. conducted an insightful investigation into the evaluation criteria of VOCALOID music compositions. (Sasaki et al., 2012) Their research entailed a comparative analysis of the interrelationships amongst variables such as the number of views, my-lists, and comments, along with their corresponding ratios. Furthermore, they posited the cogency of incorporating these parameters into the evaluation criteria. It is notable that this model bears a significant resemblance to the approach we adopted in our data sampling process.

In the same year, Yamagishi and Saito embarked on an examination of the popularity of VOCALOID tracks, focusing specifically on how attributes such as tone duration, pitch, range, and rhythm contributed to their overall popularity. (Yamagishi & Saitou, 2012). Consistent with our hypothesis, their findings inferred that the VOCALOID's capacity to perform songs challenging for real human vocalization largely accounts for its popularity amongst listeners. Notably, our data sampling method aligns with theirs, employing the same source (*Weekly VOCAL Character Ranking*).

Further down the timeline, in 2015, Yokoyama and Saitou analyzed the popularity of J-POP music spanning from 1980 to 2007 (Yokoyama & Saitou, 2015), again in connection to its characteristic features, which include chord progressions, rhythms, and motives. By comparing the proportions of these elements annually, they concluded that highly popular J-POP increasingly exhibited convergence year after year.

Most recently, Nakai and Kamimura (2022) compared VOCALOID music to J-POP music in terms of lyrics (Nakai & Kamimura, 2022). They concluded that VOCALOID music used more nouns than J-pop, and had less specific meaning. This research also informed our present study's sampling methodology: due to factors such as Japan's unique idol culture (which contains a strong dance/visual component), CD and music-streaming sales charts are not the most indicative of popularity. Accordingly, the authors used the top requested karaoke songs from the previous year. In our case---as described below—we used Apple music rankings (which incorporate music video success.)

In sum, there is still much to be discovered about VOCALOID music, and given that there are no existing melodic comparisons between VOCALOID music and J-POP music, we set out to create a corpus and examine these musical features systematically.

2. Methods

The project methodology can be divided into two main components: data collection and analysis. In terms of data collection, as described below, we have different strategies for sampling VOCALOID music and J-POP music. For data analysis, we use humdrumR (Condit-Schultz & Arthur, 2019) to perform our computational musicological analyses and compare the two groups.

2.1. Materials

In this section we describe how we created a corpus of VOCALOID and J-pop music from 2007 to 2020, resulting in a total of 140 songs.

2.1.1. Data Collection for VOCALOID Music

For selecting VOCALOID songs, we collected data from Weekly VOCAL Character Ranking (Sippotan, 2007-2022), which is a set of videos submitted to a Japanese video-sharing site (Niconico—known as Nico Nico Douga before 2012) and is the primary source of most Japanese VOCALOID music for introducing the weekly lists of top songs sung with an electronic vocal synthesis engine (including Synthesizer V, UTAU, and Cevio)..

We chose this source because it is the most authoritative ranking in the VOCALOID community, and it takes into account the views, comments, my-lists, and likes, which is much more persuasive than number of views alone.

There have been several changes over time to how the rankings are scored to ensure fairness as technology changes. The current formula is: "we define the score of the song as s, the number of views as v, the number of comments as c, the number of my-lists as m, and the number of likes as l. The current scoring formula for *Weekly VOCAL Character Ranking* is:

$$s = v \times X + c \times A + m \times B + l \times C$$

Where:

$$A = (v + m) \div (v + c + m)$$

$$B = (m \div v \times 100) \times 2$$

$$C = (l \div v \times 100) \times 2$$

$$X = (B + C) \div 6$$

$$-\max(X) = 1$$

Since the songs are only published by week and we wished to select the top five songs of each year, our process was as follows: We selected the top ten songs from each week (from January 2007 through December 2020), aggregated them annually, and narrowed them down according to the greatest number of overall appearances through all weekly rankings of that year, until we chose five representative "hit" VOCALOID songs for each year. Our methodology assumes that hit songs with greater overall persistence will be the most representative of the typical VOCALOID aesthetic.

A song may appear several times in the sample list across different years based on this sampling methodology (for example, the song *Senbon Zakura* was the first in both 2011 and 2015). In this case, that song will represent the year it first appeared and will be excluded in subsequent years (for example, *Senbon Zakura* represents 2011 instead of 2015).

In 2019, the *weekly vocal character ranking* was divided into a general ranking and a New Songs ranking. We did not consider the New Songs ranking because this ranking system does not allow us to measure persistence over time.

Once we identified the full set of sampled VOCALOID songs, we manually searched for the editor files of bette versions (usually sung by Synthesizer V or UTAU) uploaded to the major VOCALOID sharing sites. After manual checking against the original song and manual error correction, they were exported as MIDI files.

The cutoff point of 2020 is due to the data collection strategy – our VOCALOID music samples consisted of cover versions, and the cover version typically happens some time after an original song's release. As such, we sampled from both groups (VOCALOID, J-pop) only until the year 2020.

Of note is that producers of cover versions prefer to use Synthesizer V and UTAU software. Their editor files could be directly output as MIDI files.

The original version of Synthesizer V was released in 2018 and was also considered part of the VOCALOID community. However, in the early days of its release, it was intended more for cover use than for composing original songs. Although weekly vocal character ranking included Synthesizer V in its statistics from early on, until 2020, Synthesizer V did not have enough influence to get a top 5. Therefore, Synthesizer V was only used as a link in the middle of the data collection and did not affect the scope of our sampling. That is, there is no one song originally sung by Synthesizer V in our data sample.

Utau is a song synthesis software that was made public in 2008. Since Synthesizer V and Utau can use the same file format, they were also considered during the data collection phase. As for the sampling scope, Utau is not counted in the same ranking as VOCALOID, Synthesizer V, and Cevio, so it did not pose any influence.

As for the tempo, the BPM information of all VOCALOID songs was embedded in the MIDI file.

Utau is a song synthesis software that was made public in 2008. Since Synthesizer V and Utau can use the same file format, they were also considered during the data collection phase. As for the sampling scope, Utau is not counted in the same ranking as VOCALOID, Synthesizer V, and Cevio, so it did not pose any influence.

As for the tempo, the BPM information of all VOCALOID songs was embedded in the MIDI file.

2.1.2. Data Collection for J-POP Music

We sampled J-POP music from Apple Music. Apple Music offers the top lists in the J-pop genre each year, called *邦楽 ヒッツ：20XX 年*, and we need 2007-2020.

We set this time range to match that of VOCALOID in order to make a fair comparison. Thus, for each year (2007-2020), we sampled the top 5 songs according to Apple's J-pop hits.

To obtain a symbolic score, we searched from available existing music scores uploaded to the MuseScore platform by users. Because of the unreliability of user uploads, we extensively reviewed and edited the MuseScore files in order to get as accurate a melodic line as possible. Importantly, many J-pop song transcriptions uploaded to MuseScore contained multiple parts. Since we required a monophonic melody, we manually isolated the primary melodic line (see Section 2.2.1).

2.2. Procedure

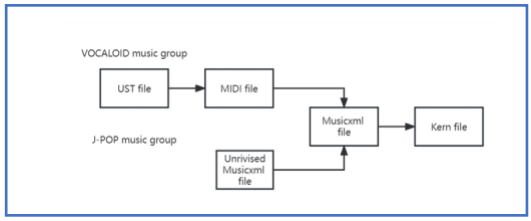

Figure 1. Format conversion

2.2.1.Data cleaning

Our process of data cleaning consisted of several tasks. For the J-pop music collection, user-uploaded versions frequently contained numerous pitch errors, incorrect keys and/or key signatures, combined harmonic accompaniment and vocal melody in a single track (i.e., musical stave), and were sometimes made as arrangements using different combinations of instruments compared with the original recording (e.g., string quartet). In addition, in this style of music, it is not uncommon for the melody to pass between different members of the group or for the group to sing in harmony, making it difficult to analyze a single melodic line. Accordingly, we manually corrected all key and

pitch information, and for each file, we aggregated all the melodic activity from the various parts into a single, newly added part (i.e., stave), ensuring that the melody was strictly monophonic and closely matched the original song. If the uploaded transcription file contained lyrical information, we preserved it.

For the VOCALOID data, as the source was UST files of user-shared song covers (usually containing only the main melody information), we similarly faced several issues related to file quality. User-uploaded versions at times contained pitch or rhythmic errors. Thus, we inspected and corrected the pitch and rhythm for each file from beginning to end. Importantly, as these are cover versions, the VOCALOID ``performers'' may differ in gender from the original artist, which typically results in key changes or octave shifts. Therefore, we also manually corrected the key and pitch range of the melody to match the original. Finally, we encountered the same problem of the main melody being split across multiple parts. In such cases, we again consolidated the main melody into a new, additional part. As with the J-pop files, if the original file contained lyrical information, we preserved it. However, unlike the J-pop group, UST files do not directly carry lyrics but rather their pronunciation (which could be Japanese kana or Romanized pronunciation). Older UST files may also directly use phonemes, rendering the lyrics information less readable in such cases.

2.2.2. Data analysis

All data analysis was conducted using the humdrumR software package (Sapp, 2023)

In order to investigate our first hypothesis (VOCALOID singers sing higher), we analyzed pitch height in three ways: the *average pitch*, *median pitch*, and *highest pitch* of the main melody in each song. First, we extracted the pitch information for each melody measured in MIDI number. Next, we calculated the median and maximum values, ignoring duration. The calculation for the average pitch, however, took duration into account, weighing each pitch by its corresponding duration, resulting in a *duration-weighted average pitch*.:

$$average\ pitch = \frac{\sum pitch \times durtaion}{\sum durtaion}$$

At this point, each song in the corpus had three corresponding pitch attributes: average pitch, median pitch, and highest pitch. Finally, we calculated the average of each of these three attributes across all pieces for each group in the corpus, resulting in the *average median pitch*, *average highest pitch*, and *average mean pitch* for the 70 pieces in each group.

To test our second hypothesis (VOCALOID singers use frequent and larger melodic intervals) we analyzed intervallic motion with two measures: the *average melodic interval* and the *maximum melodic interval* of each song's melody. The melodic intervals for of each piece were measured in semitone units.

To test our third hypothesis (VOCALOID singers sing faster), we analyzed this feature using two metrics: *BPM* (Beats Per Minute) and the *density of notes* in each melody. As mentioned, BPM was encoded in the original files (one value per file). For note density, we calculated the average number of *note onsets per measure*. Since songs have different lengths, we normalized this measure by dividing by the total number of measures in the song.

3. Results

The results of this project can be divided into two parts:

First, findings related to the previously proposed hypotheses, and second, a humdrum dataset of J-pop and VOCALOID music containing 140 files split between the two genres. (Dong & Arthur, 2023)

3.1.Hypothesis tests

In order to make a valid comparison of J-pop and VOCALOID, we had to isolate our pitch-related comparisons to that of only female artists due to the limited sample size of male vocalists within the VOCALOID group. Accordingly, the pitch findings pertaining to this subgroup lack sufficient statistical power and thus are not included in this paper.

Upon focusing exclusively on the female vocalist subset, it is observed that the duration-weighted average pitch of VOCALOID singers is slightly higher than that of J-POP singers, with the difference amounting to less than one whole tone (66.3 for J-pop and 68.0 for VOCALOID, while $t = -3.77$, $p = 0.0003$). While the difference is significant, the size of the difference is smaller than expected.

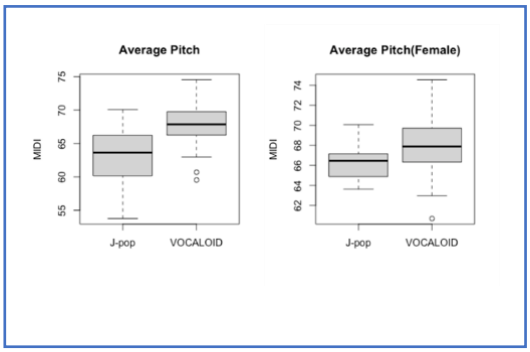

Figure 2. Duration-weighted average pitch comparison

Similarly, VOCALOID artists exhibit a marginally elevated median pitch relative to J-Pop artists, albeit again less than one whole tone (66.0 for J-Pop and 67.9 for VOCALOID). Finally, as shown in Figure 3, VOCALOID singers surpass J-Pop singers in average maximum pitch by more than a major third (75.1 for J-Pop and 79.2 for VOCALOID).

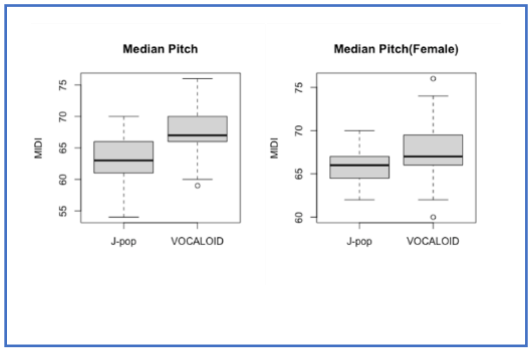

Figure 3. Median pitch comparison

Contrary to our second hypothesis, neither the proportion of skips within Vocaloid melodies nor the mean intervals demonstrate a significant difference when compared to J-Pop. Additionally, there is no significant difference in the proportion of skips. The mean proportion of skip (intervals greater than two semitones) in J-Pop amounts to 20.1%, as opposed to 19.2% in Vocaloid ($t = 1.04$, $p = 0.30$). The average interval difference for J-Pop measures 1.77, in contrast to 1.67 in Vocaloid ($t = 1.57$, $p = 0.12$). Furthermore, the average maximum interval in J-Pop is 9.56, while it is 9.76 in Vocaloid ($t = -0.59$, $p = 0.55$).

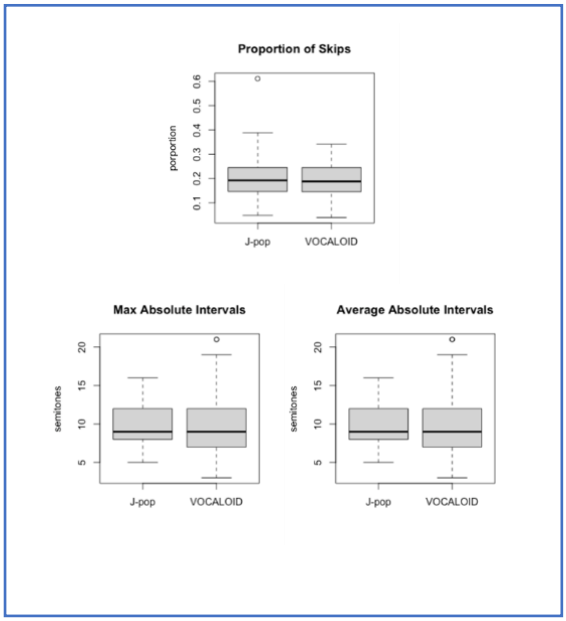

Figure 4. Skips and Pitch interval comparison

Vocaloid singers exhibit a notably high singing pace in comparison to J-Pop performers. The mean beats per minute (BPM) of Vocaloid compositions is approximately 36.5 BPM higher than that of J-Pop (117.7 BPM for J-Pop and 154.9 BPM for Vocaloid) across a representative 14-year span of tracks from 2007 to 2020. However, the note density of Vocaloid songs is not significantly different with that of J-Pop (6.2 for J-pop and 6.6 for VOCALOID, while $t = -1.36$, $p = 0.1773$).

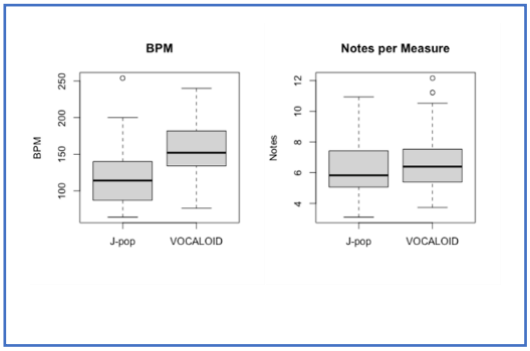

Figure 5. BPM and Notes Density comparison

Due to the limited sample size of male vocalists within the VOCALOID group, the findings pertaining to this subgroup lack sufficient statistical power and thus are not included in this paper.

Upon focusing exclusively on the female vocalist subset, it is observed that the duration-weighted average pitch of VOCALOID singers is slightly higher than that of J-POP singers, with the difference amounting to less than one whole tone (66.3 for J-pop and 68.0 for VOCALOID, while the p-value of t-test on these two groups is 0.0003). The difference is not great, but the difference does exist.

With regard to the mean median pitch evaluated on a per-file basis, VOCALOID vocalists exhibit a marginally elevated pitch relative to J-Pop artists, albeit less than one whole tone (66.0 for J-Pop and 67.9 for VOCALOID).

As figure 10 shows, A more significant discrepancy is observed in the highest pitches, wherein VOCALOID singers surpass J-Pop singers by a magnitude exceeding two whole tones (75.1 for J-Pop and 79.2 for VOCALOID).

Contrary to our initial hypothesis, neither the proportion of skips within Vocaloid melodies nor the mean intervals demonstrate a significant difference when compared to J-Pop. Additionally, there is no substantial disparity in the average maximum interval ratios on a per-file basis. The mean proportion of skips in J-Pop amounts to 20.1%, as opposed to 19.2% in Vocaloid, while the p-value of t-test on these two groups is 0.3007. The average interval difference for J-Pop measures 1.77, in contrast to 1.67 in Vocaloid. Furthermore, the average maximum interval in J-Pop is 9.56, while it is 9.76 in Vocaloid.

Vocaloid singers exhibit a notably high singing pace in comparison to J-Pop performers. The mean beats per minute (BPM) of Vocaloid compositions is approximately 36.5 BPM higher than that of J-Pop (117.7 BPM for J-Pop and 154.9 BPM for Vocaloid) across a representative 14-year span of tracks from 2007 to 2020. Additionally, the note density of Vocaloid songs is not significantly

different with that of J-Pop (6.2 for J-pop and 6.6 for VOCALOID, while the p-value of t-tests on these two groups is 0.1773).

3.2. Caveats

Data collection based on user uploads inherently encompasses numerous errors. Despite rigorous attempts to rectify them, imperfections persist.

The two groups (VOCALOID and J-POP) exhibit a striking disparity in the proportion of vocals by gender. The VOCALOID group is predominantly characterized by a female vocal, whereas the J-POP group is predominantly male. The variance in pitch between male and female vocals necessitates the consideration of female-only vocals in this study when comparing pitches, primarily due to the dearth of male vocal data in the VOCALOID group. Consequently, a holistic pitch comparison between the two groups, excluding gender consideration, offers limited insights. Unfortunately, the sample size is reduced analysis is confined to female vocals.

Furthermore, this project derived the speed of singing by considering the BPM together with the average number of notes per measure, which is not as intuitive and precise as converting it to a duration-based unit, for example, number of notes per second.

4. Discussion

This paper outlined the creation of a new corpus of VOCALOID and J-pop music, and presented the results of a comparative musicological analysis according to several hypotheses relating to the casual observations in distinctions between the two styles, namely, differences in pitch height, use of melodic intervals and tempi. While some modest differences in pitch and tempi were observed (though not in the use of melodic intervals), we have merely scratched the surface in this analysis. Future work might explore other musicological or performance aspects not analyzed here, such as timbre, texture, song structure, or lyrics.

References

Donna Li. (2021). The Rise of Vocaloid, a Voice Synthesizing Software That Has Revolutionized the Digital Music Industry. *The Science Survey*. https://thesciencesurvey.com/arts-entertainment/2021/06/22/the-rise-of-vocaloid-a-voice-synthesizing-software-that-has-revolutionized-the-digital-music-industry/

Billboard Japan. (2019). *'Lemon' Singer Kenshi Yonezu Dominated J-Pop in 2019: Here's Why*. Billboard Music News.

Finan, D. (2020). Interrogating the global success of J-pop and K-pop idols. https://doi.org/10.1386/eapc_00023_1

Michael Bourdaghs. (2012). Sayonara Amerika, Sayonara Nippon, A Geopolitical Prehistory of J-Pop. Columbia University Press.

Asuka Sasaki, Shinich Nakayama, Tetsuya Maeshiro. (2012). Analysis of the relationship between lyrics and melody in popular Vocaloid songs. The 75th National Convention Lecture and Paper Series, pp 843-844.

Yuki Yamagishi, Kazumi Saitou. (2012). Evaluation Methods and a Recommendation Method for VOCALOID Music. Research Report Music Information Science (MUS) 2012 (2), pp 1-6.

Masao Yokoyama, Youya Saitou. (2015). Analysis of Japanese pop music ranked the Top 10 in Oricon chart. https://doi.org/10.1525/mp.2005.22.4.663

Yuka Nakai, Aaiki Kamimura. (2022). Factors affecting the popularity of Vocaloid songs: comparison with J-POP songs based on word class analysis. https://doi.org/10.1525/mp.2005.22.4.663

N. Condit-Schultz and C. Arthur, "humdrumR: A New Take on an Old Approach to Computational Musicology," Proceedings of the International Society of Music Information Retrieval Conference., pp. 715–722, 2019.

Sippotan. (2007-2022). *Weekly VOCAL Character Ranking*. https://www.nicovideo.jp/series/85093.

C. Sapp. (2023). Humlib Library Accessed. [Online]. Available: https://humlib.humdrum.org/

Arthur Claire, Bowen Dong(2023) JPOP-VOCALOID Dataset. [Online]. Available: https://github.com/catallerG/JPOP-VOCALOID