Lyrics Information Processing: Analysis, Generation, and Applications

Kento Watanabe and **Masataka Goto**
National Institute of Advanced Industrial Science and Technology (AIST)
{kento.watanabe, m.goto}@aist.go.jp

Abstract

In this paper we propose *lyrics information processing (LIP)* as a research field for technologies focusing on lyrics text, which has both linguistic and musical characteristics. This field could bridge the natural language processing field and the music information retrieval field, leverage technologies developed in those fields, and bring challenges that encourage the development of new technologies. We introduce three main approaches in LIP, 1) lyrics analysis, 2) lyrics generation and writing support, and 3) lyrics-centered applications, and briefly discuss their importance, current approaches, and limitations.

1 Introduction

For songs that are musical pieces with singing voices, lyrics text is one of key factors that make listeners feel songs are attractive because it delivers messages and expresses emotion. Since the lyrics text plays an important role in music listening and creation, some studies in the music information retrieval (MIR) community have already focused on it, but not as many as studies that have focused on musical audio signals and musical scores. Similarly, in the natural language processing (NLP) community there have not been many studies focusing on lyrics text, and most NLP methods assume prose text, not lyrics text. Since lyrics text is a series of words, some NLP methods could be applied to it successfully, but NLP methods are not always effective for lyrics text because the natures of lyrics and prose texts are different as described in Section 2.

We therefore propose to refer to a broad range of lyrics-related studies as *lyrics information processing (LIP)*, which could also be considered music information processing for lyrics texts. LIP shares some core technologies with NLP and MIR, and research and development of LIP could contribute to the MIR and NLP communities as follows:

(1) Academic contributions: Since lyrics are an important aspect of music information, LIP could broaden the scope of MIR and complement it. Since lyrics are a difficult form of natural language, LIP could provide challenging issues that are not addressed by existing NLP technologies. The nature of lyrics (e.g., style, structure, and semantics) could also be investigated by automatically analyzing and generating lyrics text data.

(2) Industrial contributions: LIP could open up practical applications that are useful for listeners and creators, such as lyrics classification, lyrics exploration, lyrics summarization, and lyrics writing support.

This paper gives an overview of LIP by categorizing lyrics-related studies into three main approaches: lyrics analysis, lyrics generation, and applications. Since the concept of LIP is broad and still emerging, we hope that this paper could stimulate further development of LIP.

2 Lyrics analysis

Because lyrics and poetry[1] have unique linguistic properties, NLP technologies for prose text are not always effective enough to analyze lyrics text. In this section we introduce studies of lyrics analysis regarding the structure and semantics of lyrics and its relationship with audio.

2.1 Lyrics structure analysis

Rhyme scheme identification: The rhyme scheme is the pattern of rhymes at the end of lyric lines. It is usually represented by using a series of letters corresponding to lines, in which repeated letters indicate rhymed lines. In the following example (RWC-MDB-P-2001 No.83 (Goto et al., 2002)),

[1]Lyrics and poetry are different types of text because lyrics are assumed to be sung along with music. However, some linguistic properties of lyrics and poetry overlap.

two consecutive lines having the same letter rhyme:

A: *Race the clock I got to score*
A: *Work or play Back for more*
B: *Only true believers rise to the top,*
B: *Licking the cream of the crop*

This rhyme scheme is "AABB" and is called *Couplet*[2]. Since typical prose text analyzers such as part-of-speech analyzers and grammar tree parsers cannot analyze rhyme schemes, some studies addressed the rhyme scheme identification task. Given a few lines of lyrics (paragraph or stanza) as the input, their rhyme scheme (ABC label sequence) is estimated. For example, Reddy and Knight (2011) and Addanki and Wu (2013) estimated the rhyme scheme by using language-independent unsupervised methods (e.g., hidden Markov models) that do not depend on morphological and phonological properties.

Lyrics segmentation: While the rhyme scheme is a line-by-line repetitive structure, lyrics also have a paragraph-by-paragraph structure like verse-bridge-chorus. Paragraphs are usually separated by a blank line, but in some lyrics they are not. Some studies therefore tackled the lyrics segmentation task in which the boundaries between paragraphs are estimated from lyrics without blank lines (Watanabe et al., 2016; Fell et al., 2018). They showed that the self-similarity matrix, which is often used in music structure analysis of audio signals in the MIR community, can be applied to lyrics text to improve the performance of lyrics segmentation. This is a good example of integrating NLP and MIR methods to accomplish a LIP task.

Verse-bridge-chorus labeling: Given paragraphs of lyrics, assigning a structural label such as verse, bridge, and chorus to each paragraph is also an important task. Simple rule-based methods such as a method of grouping paragraphs with the same label (Baratè et al., 2013) and a method of labeling each paragraph (Mahedero et al., 2005) have been proposed. Since a sufficient amount of lyrics data annotated with structural labels is still lacking for machine-learning approaches, there is much room for improvement.

2.2 Lyrics semantic analysis

Emotional expressions, topics, and stories in lyrics are factors that have a great influence on listeners' emotions. Since lyrics tend to be constrained by melody lines and have a limited length, a typical way of expressing messages in lyrics is different from the way they are expressed in prose text. Lyrics messages are often emotional, inspiring, concise, and (intentionally) obscure. Even if detailed moods, topics, and stories are not explicitly described in lyrics, listeners can enjoy guessing or inferring them. Some studies have already analyzed such semantic factors behind lyrics text.

Mood estimation: Supervised learning-based methods estimating the mood or emotion of lyrics have been developed (Wang et al., 2011; Hu and Downie, 2010; Delbouys et al., 2018) and are based on a word dictionary in which valence and arousal values (Russell, 2003) are annotated (Bradley and Lang, 1999; Warriner et al., 2013). Since a lot of mood estimation methods for audio signals have been proposed in the MIR community, it would be possible to develop mood estimation based on both lyrics text and audio. In the future, unsupervised methods and support for low-resource languages are expected to be developed because supervised learning-based methods require training data of annotated lyrics, which are language-dependent.

Topic modeling: For lyrics topic modeling, unsupervised methods such as latent Dirichlet allocation (LDA), non-negative matrix factorization, and their extensions are often used (Kleedorfer et al., 2008; Sasaki et al., 2014; Tsukuda et al., 2017). Unlike mood estimation methods, these methods do not require training data with valence and arousal values, which results in the advantage of easily preparing training data for different languages. The obtained word topics (clusters) are further used as clues for classification tasks or used in visualization functions for music exploration. It is, however, difficult to appropriately evaluate the accuracy of topics obtained by unsupervised learning. A previous study tackled this difficulty by evaluating the correlation between estimated topics clusters and human-annotated ones (Sterckx et al., 2014).

Storyline modeling: Lyric writers consider themes and stories when writing lyrics. For the verse-bridge-chorus structure of lyrics, an example of a storyline represented as a topic transition is *introduction* (verse) $\rightarrow$ *past event* (bridge) $\rightarrow$ *emotional message* (chorus). Watanabe et al. (2018b) proposed an extended hidden Markov model to learn this topic transition structure from lyrics data without supervision. Their model learned topic transitions that are often found in love songs, hip-

[2]There are various rhyme schemes, such as ABAB (*Alternate Rhyme*), ABABBCBC (*Ballade*), AAAAA (*Monorhyme*), AAABBB (*Triplet*), and ABBA (*Enclosed Rhyme*).

hop songs, and so on even if they are not explicitly given.

2.3 Analysis of the relationship between lyrics text and music audio

A clear difference between lyrics and poetry is the presence or absence of accompanying music. Since investigating the relationship and synchronization between lyrics and music audio is an important topic of research, there have been various related studies that deal with the relationship between syllable stress and pitch (Nichols et al., 2009), the relationship between words and chords (Greer et al., 2019), the relationship between rests in melody and boundaries of words, lines, and paragraphs (Watanabe et al., 2018a), and lyrics-to-audio alignment (Kan et al., 2008; Fujihara et al., 2011; Mauch et al., 2012; Chien et al., 2016; Chang and Lee, 2017; Stoller et al., 2019; Gupta et al., 2019).

3 Lyrics generation and writing support

As natural language generation (NLG) has been actively researched, automatic lyrics generation is becoming a popular topic of research. NLG technologies have been greatly improved in performance by deep neural networks (DNNs) and are utilized in applications such as machine translation and dialogue systems. Generating poetry and novels has also been developed, though generating creative text is challenging. Generating lyrics is also challenging and has further technical difficulties caused by lyrics-specific musical constraints such as melodies and rhymes. In this section we introduce studies of lyrics generation as well as writing support systems that utilize lyrics generation methods.

3.1 Automatic lyrics generation

Rhyme-scheme-conditioned lyrics generation: Since lyrics and poetry often have rhyme schemes as introduced in Section 2.1, some studies have addressed the task of generating lyrics and poetry that satisfy constraints of a rhyme scheme (Barbieri et al., 2012; Hopkins and Kiela, 2017). In automatically generating lyrics, most methods use language models such as n-grams and recurrent neural networks as well as word sequence search based on the Markov process. To deal with the constraints, several extended word-sequence search methods have been proposed, such as those using the strong constraint that words that do not satisfy the rhyme scheme are discarded during word sequence search and the weak constraint that the score is calculated based on how well the given rhyme scheme is satisfied.

Melody-conditioned lyrics generation: Although most studies of automatic lyrics generation have generated lyrics using only text data without considering musical audio signals and musical scores, some studies have addressed the task of generating fluent lyrics that are singable when a melody (a sequence of musical notes) is given (Lu et al., 2019). Watanabe et al. (2018a) confirmed that the frequency of word/line/paragraph boundaries depends on the duration of rests and proposed an advanced lyrics language model that takes advantage of this dependency. Their method can generate segmented lyrics that are singable for the verse-bridge-chorus structure of the input melody. It, however, requires training data in which lyrics syllables and melody notes are aligned. Such data could be easily created if technologies such as the above-mentioned lyrics-to-audio alignment, lyrics recognition (Hosoya et al., 2005; Dabike and Barker, 2019; Suzuki et al., 2019), and melody note transcription (Yang et al., 2017; Román et al., 2018; Nishikimi et al., 2019) could mature in the future.

Automatic generation of structured lyrics: Most lyrics generation systems can generate only one paragraph of lyrics, though lyrics have some paragraphs in general. This is because language models for lyrics did not explicitly capture the consistency of topics and relations between paragraphs. Watanabe et al. (2014) have proposed a probabilistic model that captures topic transitions between paragraphs to generate lyrics having the storyline. Fan et al. (2019) have proposed a lyrics generation method using the long short-term memory language model that captures the hierarchical structure of words, lines, and paragraphs to leverage the dependency of long word sequences. Although these studies have made it possible to generate lyrics that are almost consistent in topic, it is still difficult to generate lyrics that are consistent in meaning.

Ghostwriting: Ghostwriting is a task of generating new lyrics that follow the style (e.g., rhyme scheme, phrasing, content, and the number of words per line) of a given artist. Potash et al. (2015) proposed a rap-lyrics generation method based on data-driven learning of the artist's style using a DNN-based language model trained with the artist's lyrics corpus.

3.2 Writing support system with automatic lyrics generation

Automatic lyrics generation makes it possible to develop systems that support lyrics writing. It is not easy for novices to write lyrics by thinking of appropriate words and phrases while considering various constraints and properties. Since candidate word sequences satisfying various constraints can be generated automatically, it is useful to show them to lyric writers to support their creative activities. Some studies have developed interactive systems that support lyrics writing by repeatedly recommending candidate word sequences that satisfy constraint parameters input by the user.

pâtissier (Abe and Ito, 2012) is an interface that allows the user to specify syllable counts, syllable stress, and vowels, and generates candidate sentences that satisfy them. *DeepBeat* (Malmi et al., 2016) is an interface that generates and suggests next-line candidates that rhyme with a line entered by the user. *LyriSys* (Watanabe et al., 2017) and *Co-PoeTryMe* (Oliveira et al., 2019) are interfaces that allow the user to specify song structure and syllable counts, select or enter topics and keywords for each paragraph, and make the system generate candidate lyrics that satisfy them. These interfaces also allow the user to manually edit the generated lyrics.

4 Applications for a collection of lyrics

Like NLP technologies, LIP technologies are useful in developing various applications, such as classification, exploration, and summarization, for a large collection of lyrics data.

4.1 Lyrics classification

Given a collection of lyrics, it is useful to classify and visualize them. Genre classification for lyrics is a popular approach that has already been studied (Mayer et al., 2008; Mayer and Rauber, 2011; Tsaptsinos, 2017). Some characteristics peculiar to lyrics (e.g., rhyme scheme, structure, meaning, and relationship with audio) have been used as features to train a supervised classifier.

4.2 Lyrics exploration

If a user wants to see the lyrics of a song the user knows, simple text-based lyrics retrieval is enough, but if a user wants to encounter unfamiliar but interesting lyrics, a content-based music exploration system focusing on lyrics is necessary. Baur et al. (2010), Sasaki et al. (2014), and Tsukuda et al. (2017) have developed such exploration systems that visualize topics of lyrics and similar artists by analyzing the content of lyrics using LDA, self-organizing maps, and so on. *Query-by-Blending* (Watanabe and Goto, 2019) is a music exploration system that enables a user to give flexible queries related to lyrics, audio signals, and artist tags by using a unified latent vector space with these three different modalities embedded.

4.3 Lyrics summarization

In browsing a collection of lyrics, a short summary of lyrics of each song helps navigate quickly. Fell et al. (2019) improved the performance of the lyrics summarization task by combining a general document summarization method with an audio thumbnailing method. Summarization more advanced than simply extracting lines, such as phrase paraphrasing and compression, requires development of advanced technologies for lyrics semantic analysis.

5 Conclusion

In this paper we have provided an overview of lyrics information processing (LIP) and have described examples of studies from the viewpoint of lyrics analysis, lyrics generation, and applications. Those examples are just excerpts taken from a variety of previous studies and possible future technologies. For example, the limited space does not allow us to discuss the relationship with *singing information processing (SIP)* (Goto et al., 2010; Goto, 2014; Humphrey et al., 2019), though we mentioned the lyric-to-audio alignment. Since lyrics are sung by singers, there are many possibilities to investigate the relationship between lyrics and the corresponding singing expressions and styles. Lyrics are thus linguistic, musical, and singable from the NLP, MIR, and SIP viewpoints, respectively. Since LIP is an emerging interdisciplinary research field that could be related to various technologies and disciplines such as natural language processing, music information retrieval, machine learning, human-computer interaction, visualization, signal processing, linguistics, and musicology, we expect research on LIP to progress in coming years from a diverse viewpoint by attracting more attention due to its importance and potential.

6 Acknowledgments

This work was supported in part by JST ACCEL Grant Number JPMJAC1602, and JSPS KAKENHI Grant Number 20K19878 Japan.

References

Chihiro Abe and Akinori Ito. 2012. A Japanese lyrics writing support system for amateur songwriters. In *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2012)*, pages 1–4.

Karteek Addanki and Dekai Wu. 2013. Unsupervised rhyme scheme identification in hip hop lyrics using hidden markov models. In *Proceedings of The First International Conference of the Statistical Language and Speech Processing (SLSP 2013)*, volume 7978, pages 39–50.

Adriano Baratè, Luca A. Ludovico, and Enrica Santucci. 2013. A semantics-driven approach to lyrics segmentation. In *Proceedings of the 8th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP 2013)*, pages 73–79.

Gabriele Barbieri, François Pachet, Pierre Roy, and Mirko Degli Esposti. 2012. Markov constraints for generating lyrics with style. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, volume 242, pages 115–120.

Dominikus Baur, Bartholomäus Steinmayr, and Andreas Butz. 2010. SongWords: Exploring music collections through lyrics. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 531–536.

Margaret M. Bradley and Peter J. Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology.

Sungkyun Chang and Kyogu Lee. 2017. Lyrics-to-audio alignment by unsupervised discovery of repetitive patterns in vowel acoustics. *IEEE Access*, 5:16635–16648.

Yu-Ren Chien, Hsin-Min Wang, and Shyh-Kang Jeng. 2016. Alignment of lyrics with accompanied singing audio based on acoustic-phonetic vowel likelihood modeling. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(11):1998–2008.

Gerardo Roa Dabike and Jon Barker. 2019. Automatic lyric transcription from karaoke vocal tracks: Resources and a baseline system. In *Proceedings of the 20th Annual Conference of the International Speech Communication Association (Interspeech 2019)*, pages 579–583.

Rémi Delbouys, Romain Hennequin, Francesco Piccoli, Jimena Royo-Letelier, and Manuel Moussallam. 2018. Music mood detection based on audio and lyrics with deep neural net. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 370–375.

Haoshen Fan, Jie Wang, Bojin Zhuang, Shaojun Wang, and Jing Xiao. 2019. A hierarchical attention based seq2seq model for Chinese lyrics generation. In *Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2019)*, pages 279–288.

Michael Fell, Elena Cabrio, Fabien Gandon, and Alain Giboin. 2019. Song lyrics summarization inspired by audio thumbnailing. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 328–337.

Michael Fell, Yaroslav Nechaev, Elena Cabrio, and Fabien Gandon. 2018. Lyrics segmentation: Textual macrostructure detection using convolutions. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 2044–2054.

Hiromasa Fujihara, Masataka Goto, Jun Ogata, and Hiroshi G. Okuno. 2011. LyricSynchronizer: Automatic synchronization system between musical audio signals and lyrics. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1252–1261.

Masataka Goto. 2014. Singing information processing. In *Proceedings of the 12th IEEE International Conference on Signal Processing (IEEE ICSP 2014)*, pages 2431–2438.

Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. 2002. RWC Music Database: Popular, classical, and jazz music databases. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, pages 287–288.

Masataka Goto, Takeshi Saitou, Tomoyasu Nakano, and Hiromasa Fujihara. 2010. Singing information processing based on singing voice modeling. In *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2010)*, pages 5506–5509.

Timothy Greer, Karan Singla, Benjamin Ma, and Shrikanth S. Narayanan. 2019. Learning shared vector representations of lyrics and chords in music. In *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2019)*, pages 3951–3955.

Chitralekha Gupta, Emre Yilmaz, and Haizhou Li. 2019. Acoustic modeling for automatic lyrics-to-audio alignment. In *Proceedings of the 20th Annual Conference of the International Speech Communication Association (Interspeech 2019)*, pages 2040–2044.

Jack Hopkins and Douwe Kiela. 2017. Automatically generating rhythmic verse with neural networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, pages 168–178.

Toru Hosoya, Motoyuki Suzuki, Akinori Ito, and Shozo Makino. 2005. Lyrics recognition from a singing voice based on finite state automaton for music information retrieval. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pages 532–535.

Xiao Hu and J. Stephen Downie. 2010. When lyrics outperform audio for music mood classification: A feature analysis. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 619–624.

E. J. Humphrey, S. Reddy, P. Seetharaman, A. Kumar, R. M. Bittner, A. Demetriou, S. Gulati, A. Jansson, T. Jehan, B. Lehner, A. Krupse, and L. Yang. 2019. An introduction to signal processing for singing-voice analysis: High notes in the effort to automate the understanding of vocals in music. *IEEE Signal Processing Magazine*, 36(1):82–94.

Min-Yen Kan, Ye Wang, Denny Iskandar, Tin Lay Nwe, and Arun Shenoy. 2008. LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):338–349.

Florian Kleedorfer, Peter Knees, and Tim Pohle. 2008. Oh Oh Oh Whoah! towards automatic topic detection in song lyrics. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 287–292.

Xu Lu, Jie Wang, Bojin Zhuang, Shaojun Wang, and Jing Xiao. 2019. A syllable-structured, contextually-based conditionally generation of Chinese lyrics. In *Proceedings of the 16th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2019)*, pages 257–265.

Jose P. G. Mahedero, Alvaro Martinez, Pedro Cano, Markus Koppenberger, and Fabien Gouyon. 2005. Natural language processing of lyrics. In *Proceedings of the 13th ACM International Conference on Multimedia (ACM Multimedia 2005)*, pages 475–478.

Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis. 2016. DopeLearning: A computational approach to rap lyrics generation. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 195–204.

Matthias Mauch, Hiromasa Fujihara, and Masataka Goto. 2012. Integrating additional chord information into HMM-based lyrics-to-audio alignment. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):200–210.

Rudolf Mayer, Robert Neumayer, and Andreas Rauber. 2008. Rhyme and style features for musical genre classification by song lyrics. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 337–342.

Rudolf Mayer and Andreas Rauber. 2011. Music genre classification by ensembles of audio and lyrics features. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pages 675–680.

Eric Nichols, Dan Morris, Sumit Basu, and Christopher Raphael. 2009. Relationships between lyrics and melody in popular music. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 471–476.

Ryo Nishikimi, Eita Nakamura, Masataka Goto, and Kazuyoshi Yoshii. 2019. End-to-end melody note transcription based on a beat-synchronous attention mechanism. In *Proceedings of the 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE WASPAA 2019)*, pages 26–30.

Hugo Gonçalo Oliveira, Tiago Mendes, Ana Boavida, Ai Nakamura, and Margareta Ackerman. 2019. Co-PoeTryMe: Interactive poetry generation. *Cognitive Systems Research*, 54:199–216.

Peter Potash, Alexey Romanov, and Anna Rumshisky. 2015. GhostWriter: Using an LSTM for automatic rap lyric generation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pages 1919–1924.

Sravana Reddy and Kevin Knight. 2011. Unsupervised discovery of rhyme schemes. In *Proceedings of The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011)*, pages 77–82.

Miguel A. Román, Antonio Pertusa, and Jorge Calvo-Zaragoza. 2018. An end-to-end framework for audio-to-score music transcription on monophonic excerpts. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018)*, pages 34–41.

James A. Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Shoto Sasaki, Kazuyoshi Yoshii, Tomoyasu Nakano, Masataka Goto, and Shigeo Morishima. 2014. LyricsRadar: A lyrics retrieval system based on latent topics of lyrics. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 585–590.

Lucas Sterckx, Thomas Demeester, Johannes Deleu, Laurent Mertens, and Chris Develder. 2014. Assessing quality of unsupervised topics in song lyrics. In *Proceedings of the 36th European Conference on IR Research (ECIR 2014)*, volume 8416, pages 547–552.

Daniel Stoller, Simon Durand, and Sebastian Ewert. 2019. End-to-end lyrics alignment for polyphonic music using an audio-to-character recognition model. In *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2019)*, pages 181–185.

Motoyuki Suzuki, Sho Tomita, and Tomoki Morita. 2019. Lyrics recognition from singing voice focused on correspondence between voice and notes. In *Proceedings of the 20th Annual Conference of the International Speech Communication Association (Interspeech 2019)*, pages 3238–3241.

Alexandros Tsaptsinos. 2017. Lyrics-based music genre classification using a hierarchical attention network. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 694–701.

Kosetsu Tsukuda, Keisuke Ishida, and Masataka Goto. 2017. Lyric Jumper: A lyrics-based music exploratory web service by modeling lyrics generative process. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, pages 544–551.

Xing Wang, Xiaoou Chen, Deshun Yang, and Yuqian Wu. 2011. Music emotion classification of Chinese songs based on lyrics using TF*IDF and rhyme. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pages 765–770.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4):1191–1207.

Kento Watanabe and Masataka Goto. 2019. Query-by-Blending: A music exploration system blending latent vector representations of lyric word, song audio, and artist. In *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, pages 144–151.

Kento Watanabe, Yuichiroh Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano. 2018a. A melody-conditioned lyrics language model. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2018)*, pages 163–172.

Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, Satoru Fukayama, Tomoyasu Nakano, and Masataka Goto. 2018b. Modeling storylines in lyrics. *IEICE Transactions on Information and Systems*, E101-D(4):1167–1179.

Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, and Masataka Goto. 2014. Modeling structural topic transitions for automatic lyrics generation. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation (PACLIC 2014)*, pages 422–431.

Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, Tomoyasu Nakano, Satoru Fukayama, and Masataka Goto. 2017. LyriSys: An interactive support system for writing lyrics based on topic transition. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (ACM IUI 2017)*, pages 559–563.

Kento Watanabe, Yuichiroh Matsubayashi, Naho Orita, Naoaki Okazaki, Kentaro Inui, Satoru Fukayama, Tomoyasu Nakano, Jordan B. L. Smith, and Masataka Goto. 2016. Modeling discourse segments in lyrics using repeated patterns. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, pages 1959–1969.

Luwei Yang, Akira Maezawa, Jordan B. L. Smith, and Elaine Chew. 2017. Probabilistic transcription of sung melody using a pitch dynamic model. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP 2017)*, pages 301–305.