RELATIONSHIPS BETWEEN LYRICS AND MELODY
IN POPULAR MUSIC

Eric Nichols[1], Dan Morris[2], Sumit Basu[2], and Christopher Raphael[1]

[1]Indiana University
Bloomington, IN, USA
{epnichol,craphael}@indiana.edu

[2]Microsoft Research
Redmond, WA, USA
{dan,sumitb}@microsoft.com

ABSTRACT

Composers of popular music weave lyrics, melody, and instrumentation together to create a consistent and compelling emotional scene. The relationships among these elements are critical to musical communication, and understanding the statistics behind these relationships can contribute to numerous problems in music information retrieval and creativity support. In this paper, we present the results of an observational study on a large symbolic database of popular music; our results identify several patterns in the relationship between lyrics and melody.

1. INTRODUCTION

Popular music uses several streams of information to create an emotionally engaging experience for the listener. Lyrics, melody, chords, dynamics, instrumentation, and other aspects of a song operate in tandem to produce a compelling musical percept. Extensive previous work has explored each of these elements in isolation, and certain relationships among these components – for example, the relationship between melody and chords – have also been addressed in the research community. However, despite their salience and central role in music cognition, lyrics have not been addressed by computational analysis to the same degree as other aspects of popular music.

In this study, we examine the relationship between lyrics and melody in popular music. Specifically, we investigate the assumption that songwriters tend to align low-level features of a song's text with musical features. Composer Stephen Sondheim, for example, has commented that he selects rhythms in music to match the natural inflections of speech [1], and popular books on songwriting suggest considering the natural rhythms of speech when writing melodies [2]. With this qualitative evidence in mind, we quantitatively examine relationships between text and music using a corpus of several hundred popular songs. Specifically, we investigate the general hypothesis that textual salience is correlated with musical salience, by extracting features representative of each and exploring correlations among those features.

This study contributes fundamental statistics to musi-

cology and music-cognition research, and makes the following specific contributions to the music information retrieval community:

1) We establish new features in the hybrid space of lyrics and melody, which may contribute to musical information and genre analysis as well as music recommendation.

2) We demonstrate a quantitative correlation between lyrical and melodic features, motivating their use in composition-support tools which help composers work with music and text.

3) We strengthen the connection between MIR and speech research; the features presented here are closely related to natural patterns in speech rhythm and prosody.

4) We make analysis of lyrics and melody in popular music more accessible to the community, by releasing the parsing and preprocessing code developed for this work.

2. RELATED WORK

Previous work in the linguistics and speech communities has demonstrated that inherent rhythms are present even in non-musical speech (e.g. [3,4]). Additional work has shown that the rhythms inherent to a composer's native language can influence instrumental melodic composition. Patel and Daniele [5] show a significant influence of native language (either English or French) on composers' choice of rhythmic patterns, and Patel et al. [6] extend this work to show a similar influence of native language on the selection of pitch intervals. This work does not involve text per se, only the latent effect of language on instrumental classical music. Beyond the rhythmic aspects of speech, additional work has demonstrated that vowels have different intrinsic pitches [7], and even that phonemes present in musical lyrics can influence a listener's perception of pitch intervals [8]. This work supports our claim that there is a strong connection between not only rhythmic aspects of speech and music, but also between linguistic, phonemic, pitch, and timbral aspects of speech and music.

In addition to these explorations into the fundamental properties of speech and lyrics, preliminary applications of the statistics of lyrics have begun to emerge for both creativity support tools and problems in music information retrieval and analysis. Proposing a creativity support tool to explore alignments of melodies and lyrics, [9] uses

a series of hand-coded heuristics to align a known set of lyrics to the rhythm of a known melody. Oliveira et al. [10] develop a preliminary system that addresses the problem of generating text to match a known rhythm; this works also includes a preliminary analysis of a small database to qualitatively validate the authors' assumptions.

Wang et al. [11] and Iskandar et al. [12] use higher-level properties of lyrical structure to improve the automatic alignment of recordings with corresponding lyrics. Lee and Cremer [13] take a similar approach to match high-level segments of lyrics to corresponding segments in a recording. Recent work in the music information retrieval community has also applied lyric analysis to problems in topic detection [14], music database browsing [15], genre classification [16], style identification [17], and emotion estimation [18]. This work motivates the present study and suggests the breadth of applications that will benefit from a deeper, quantitative understanding of the relationship between lyrics and melody.

3. METHODS

3.1 Data Sources and Preprocessing

Our database consisted of 679 popular music lead sheets in MusicXML format. 229 of our lead sheets came from a private collection; the remaining 450 came from Wikifonia.org, an online lead sheet repository. Our data spans a variety of popular genres, including pop, rock, R&B, country, Latin, and jazz, with a small sampling of folk music.

Each lead sheet in our database contains a melody, lyrics, and chords for a single song (chords were not used in the present analysis). Lyrics are bound to individual notes; i.e., no alignment step was necessary to assign lyrics to their corresponding notes. Word boundaries were provided in the MusicXML data so it was possible to determine which syllables were joined to make whole words without consulting a dictionary. Key and time signature information was also provided for each song (including any changes within a song). For all analyses presented in this paper, we ignored measures of music with a time signature other than 4/4. Lead sheets were processed to build a flat table of notes (pitch and duration) and their corresponding syllables, with repeats flattened (expanded and rewritten without repeats) to allow more straightforward analysis.

3.2 Computed Musical Features

This section describes the three features that were computed for each note in our melody data.

3.2.1 Metric Position

For each note, the "Metric Position" feature was assigned to one of five possible values based on the timing of the note's onset: **downbeat** (for notes beginning on beat 1), **half-beat** (for notes beginning on beat 3), **quarter beat** (beginning on beats 2 or 4), **eighth beat** (beginning on the "and" of any quarter beat), and **other**.

3.2.2 Melodic Peak

The "Melodic Peak" feature is set to **True** for any note with a higher pitch than the preceding and subsequent notes. It is set to **False** otherwise (including notes at the beginning and end of a song). We selected this feature because previous research has connected melodic contours to a number of features in instrumental music [19].

3.2.3 Relative Duration

For a note in song s, the "Relative Duration" feature is computed by calculating the mean duration (in beats) for all notes in s and then dividing each note's duration by the mean. Thus "Relative Duration" values greater than 1 indicate notes longer than mean duration for the associated song.

3.3 Computed Lyrical Features

This section describes the three features that were computed for each syllable in our lyric data, based on the syllable itself and/or the containing word.

We determined the pronunciation of each syllable by looking up the containing word in the CMU Pronouncing Dictionary [20], a public-domain, machine-readable English dictionary that provides phoneme and stress level information for each syllable in a word. In cases where the dictionary provided alternate pronunciations, we selected the first one with the correct number of syllables. Unknown words and words whose associated set of notes in our MusicXML data did not correspond in number to the number of syllables specified by the dictionary were removed from the data. Note that this dictionary provides pronunciation for isolated words. Stress patterns can change based on the surrounding context, so this pronunciation data is only an approximation of natural speech.

3.3.1 Syllable Stress

The CMU dictionary gives a stress level according to the following ordinal scale: **Unstressed**, **Secondary Stress**, and **Primary Stress**; each syllable was assigned one of these three values for the "Syllable Stress" feature. Secondary stress is typically assigned in words with more than two syllables, where one syllable receives some stress but is not the primary accent. For example, in the word "letterhead", the first syllable is assigned a primary stress, the second is unstressed, and the third is assigned a secondary stress.

3.3.2 Stopwords

Stopwords are very common words that carry little semantic information, such as "a", "the", and "of". Stopwords are generally ignored as "noise" in text processing

CMU Vowel	IPA (Pan-English)	Example
AH	ə	hut
UH	ʊ	hood
IH	ɪ	it
ER	ɜ	hurt
EH	ɛ	Ed
AE	æ	at
AA	ɑː	odd
IY	iː	eat
UW	uː	two
AY	aɪ	hide
AO	ɔː	ought
OW	oʊ	oat
EY	eɪ	ate
AW	aʊ	cow
OY	ɔɪ	toy

Table 1. Vowels used in our analysis (sorted by increasing average associated relative note duration – see section 4.4). In order to classify vowels as short, long, or diphthong, vowels from the CMU dictionary were translated to Pan-English IPA (International Phonetic Alphabet) symbols according to [23]. Symbols ending in a colon (:) represent long vowels; symbols containing two characters (e.g. oʊ) represent diphthongs. As is further elaborated in Section 4, we highlight that when sorted by average musical note duration, short vowels are correlated with shorter durations than long vowels and diphthongs in all cases, and with the exception of one long vowel (AO, or ɔː), diphthongs are assigned longer durations than long vowels.

systems such as search engines. There is no definitive or absolutely correct list of English stopwords; we use the monosyllabic subset of the online adaption [21] of the fairly canonical stopword list originally presented by van Rijsbergen [22]. We specifically choose the monosyllabic subset so that we are conservative in our identification of stopwords; we consider words such as "never", while perhaps too common for certain applications, to be semantically rich enough to merit treatment as non-stopwords. The "Stopword" feature is set to **True** or **False** for each monosyllabic word, and is undefined for multisyllable words.

3.3.3 Vowels

Each syllable in the dictionary may include multiple consonants, but only one vowel. We extract the vowel for each syllable; this categorical feature can take on one of 15 possible values, enumerated in Table 1.

4. RESULTS

Having established a set of features in both the melodic and lyrical spaces, we now turn our attention to exploring correlations among those features.
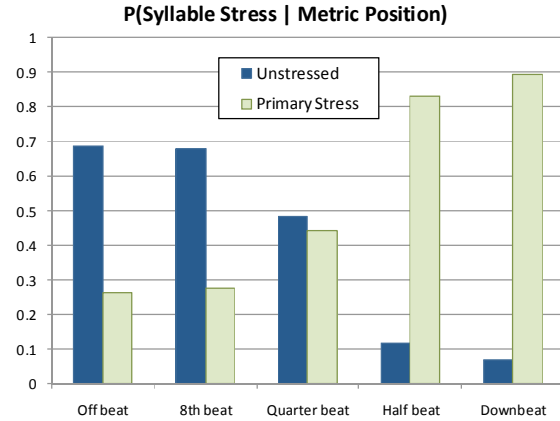
Figure 1. *P(syllable stress | metric position)*. The stronger a note's metric position, the more likely it is that the associated syllable has a primary stress. Secondary stresses are rare overall and were omitted from this graph.
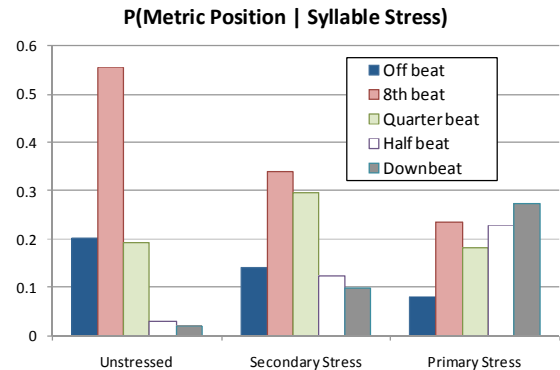
4.1 Syllable Stress

Based on our general hypothesis that musical salience is frequently associated with lyrical salience, we hypothesized that stressed syllables would tend to be associated with musically accented notes. We thus explored correlations between the "Syllable Stress" feature and each of our melodic features. Each analysis in this subsection was performed only using note data associated with polysyllabic words, so that stress values are meaningful.

4.1.1 Syllable Stress and Metric Position

A stronger syllable stress is associated with a stronger metric position, as we see in Figures 1 and 2. These give two different views of the data, based on conditioning first by either *metric position* or *syllable stress*.

Figure 1 demonstrates that the half beat and downbeat positions strongly favor stressed syllables, and are rarely associated with unstressed syllables. For comparison, stressed and unstressed syllables occur with approximately equal a priori probabilities (P(primary stress) = 0.46

Figure 2. *P(metric position | syllable stress)*. Unstressed syllables are very unlikely to show up on a downbeat, but very likely at an 8th beat position. Primary stresses rarely occur on off-beats.

P(Melodic Peak | Syllable Stress)
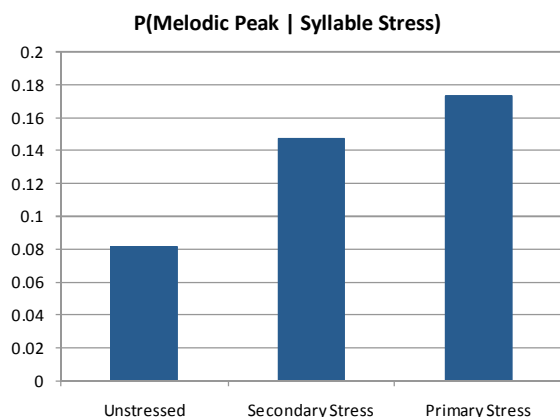
**Figure 3**. *P(melodic peak | syllable stress)*. The probability of a melodic peak increases with increasing syllable stress.

and P(unstressed) = 0.48). Figure 2 similarly shows that unstressed syllables are very unlikely to show up on a downbeat, but very likely at an 8th-beat position, and that primary stresses rarely occur on off-beats. Pearson's Chi-Square test confirms a significant relationship between these features ($p < 0.0001$).

*4.1.2 Syllable Stress and Melodic Peaks*

Figure 3 shows that stronger syllable stress is also strongly associated with the occurrence of melodic peaks. This relationship holds in both directions: the probability of a primary stress is significantly higher at syllables corresponding to melodic peaks than at non-peaks, and the probability of a melodic peak is much higher at stressed syllables than non-stressed syllables. Pearson's Chi-Square test confirms a significant relationship between these features ($p < 0.0001$).

4.1.3 Syllable Stress and Note Duration

In Figure 4, the "Relative Duration" feature has been dis-

P(Syllable Stress | Relative Duration)
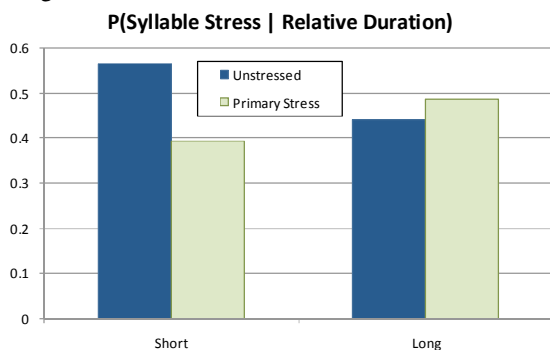
Figure 4. *P(syllable stress | relative duration)*. The "Relative Duration" feature was discretized into two values: "Short" (Relative Duration ≤ 1, i.e. notes shorter than the mean duration within a song), and "Long" (Relative Duration > 1). Shorter note durations are more likely to be associated with unstressed syllables; longer durations are more likely to be associated with stressed syllables.

P(Stopword | Metric Position)
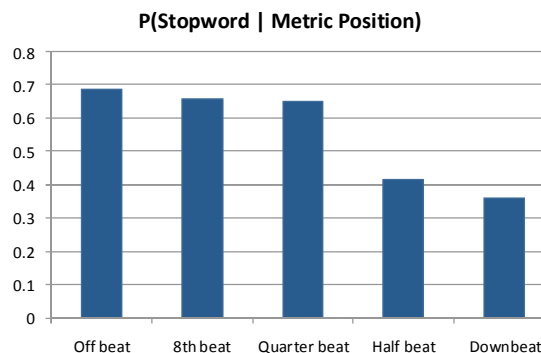
Figure 5. *P(stopword | metric position)*. This graph shows metric positions moving from weak (left) to strong (right), and the corresponding decrease in the probability of stopwords at corresponding syllables.

cretized into two values: "Short" (Relative Duration ≤ 1, i.e. notes shorter than the mean duration within a song), and "Long" (Relative Duration > 1). Figure 4 shows that long notes are more likely to associated with stressed syllables than unstressed syllables, and short notes are more likely to be associated with unstressed syllables. The inverse relationship is true as well; most notes (55%) associated with unstressed syllables are short, and most notes (55%) associated with primary-stress syllables are long. Pearson's Chi-Square test confirms a significant relationship between these features ($p < 0.0001$).
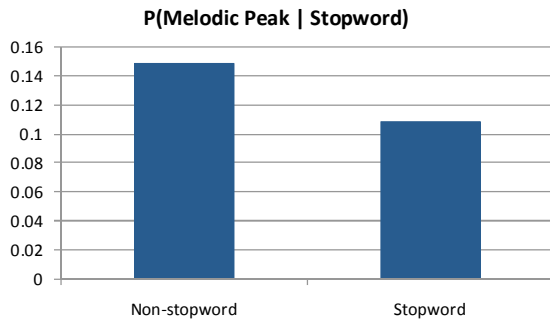
4.2 Stopwords

Based on our general hypothesis that musical salience is frequently associated with lyrical salience, we hypothesized that semantically meaningful words would tend to be associated with musically salient notes, and consequently that stopwords – which carry little semantic information – would be associated with musically non-salient notes. In this subsection, only notes associated with monosyllabic words are used in the analysis, since our list of stopwords includes only monosyllabic words.

4.2.1 Stopwords and Metric Position

Figure 5 shows the probability of finding a stopword at each metric position. The stronger the metric position, the less likely the corresponding word is to be a stopword. The overall probability of a stopword (across all metric positions) is 0.59. However, the half-beat and downbeat positions favor non-stopwords. Pearson's Chi-Square test confirms a significant relationship between these features ($p < 0.0001$).

4.2.2 Stopwords and Melodic Peaks

Figure 6 shows that melodic peaks are more frequently associated with non-stopwords than with stopwords. The inverse relationship holds as well: the probability of observing a stopword at a melodic peak is lower than at a non-peak. Pearson's Chi-Square test confirms a significant relationship between these features ($p < 0.0001$).

Figure 6. *P(melodic peak | stopword)*. Melodic peaks are significantly more likely to coincide with non-stopwords than with stopwords.

4.3 Vowels

We hypothesized that vowel sounds would vary reliably with note durations, reflecting both the aesthetic properties of different vowel types and the impact of different vowel types on a singer's performance. We thus looked at correlations between the "phonetic length" of vowels (short, long, or diphthong) and the average durations of corresponding notes. We assign phonetic length to vowel length according to the IPA convention for Pan-English interpretation of phonemes (Table 1).

4.3.1 Vowels and Relative Duration

Figure 7 is a sorted plot of mean relative duration of notes for each vowel type. In general agreement with our hypothesis, the shorter vowels all have mean relative duration less than 1 (i.e. short vowels have shorter duration than average in a song); long vowels and diphthongs have mean relative duration greater than 1 (i.e. long vowels have longer duration than average). We highlight that short vowels are correlated with shorter durations than long vowels and diphthongs in all cases, and with the exception of one long vowel (AO, or ɔ:), diphthongs are assigned longer durations than long vowels.

If we generate a Boolean feature indicating whether a vowel is long (including diphthongs) or short, and we similarly use the Boolean version of the "Relative Duration" feature (see Figure 5), we can proceed as in previous sections and correlate vowel length with relative duration. Figure 8 shows that longer notes are more likely to be associated with long vowels, and short notes with short vowels. Pearson's Chi-Square test confirms the significance of this relationship (p < 0.0001).

5. DISCUSSION

5.1 Summary of Findings

We have introduced an approach for analyzing relationships between lyrics and melody in popular music. Here we summarize the relationships presented in Section 4:

1) Level of syllabic stress is strongly correlated with strength of metric position.
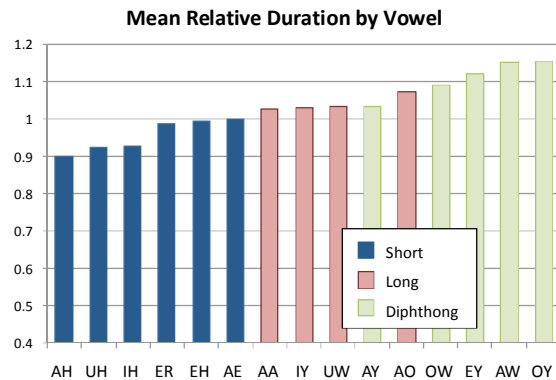
Figure 7. Mean relative duration of notes associated with each vowel, sorted form short notes (left) to long (right). The resulting partitioning of similar vowel types shows that short vowels are correlated with shorter durations than long vowels and diphthongs in all cases, and with the exception of one long vowel (AO), diphthongs are correlated with longer durations than long vowels.

2) Level of syllabic stress is strongly correlated with the probability of melodic peaks.
3) Level of syllabic stress is strongly correlated with note duration.
4) Stopwords (which carry little semantic weight) are strongly correlated with weak metric positions.
5) Stopwords are much less likely to coincide with melodic peaks than non-stopwords.
6) Short vowels tend to be associated with shorter notes than long vowels, which tend to be associated with shorter notes than diphthongs.

These findings support our highest-level hypothesis: songwriters tend to align salient notes with salient lyrics. The strength of these relationships – and our ability to find them using intuitive features in both lyrics and melody – suggests the short-term potential to apply these relationships to both MIR and creativity support tools.

5.2 Applications and Future Work

The analysis presented here used features that were easily accessible in our database of symbolic popular music. Future work will explore similar relationships among more
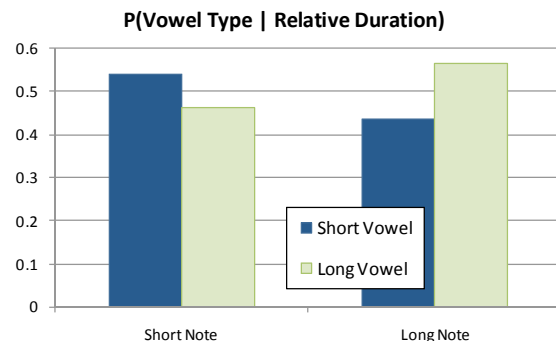
Figure 8. *P(vowel type | relative duration)*. Short notes are more frequently associated with short vowels, and long notes with long vowels.

complex features of both lyrics (e.g. valence, parts of speech) and music (e.g. tone and timbre, dynamics, and pronunciation data extracted from vocal performances).

Understanding the statistics of lyrics alone will contribute to the many of the same applications that will benefit from our understanding of the relationship between lyrics and music. Therefore, future work will also include a large-scale study that more deeply explores the statistics and grammatical patterns inherent to popular lyrics, as compared to non-musical text corpora.

Most importantly, future work will explore applications of a quantitative understanding of the relationship between lyrics and melody. For example, these relationships can provide priors for lyric transcription and lyric alignment to audio recordings. Similarly, strengthening the connection between music and lyrics will allow us to more easily borrow techniques from the speech community for problems such as artist identification and score-following for popular music.

Furthermore, a quantitative understanding of the relationship between lyrics and melody has applications in tools that support the creative process. Composers and novices alike may benefit from systems that can suggest lyrics to match a given melody or vice versa, and understanding the relationships presented in this paper is an important first step in this direction. One might similarly imagine a "grammar checker" for popular composition, which provides suggestions or identifies anomalies not in text, but in the relationship between melody and lyrics.

6. PREPROCESSING TOOLKIT

In order to stimulate research in this area and allow replication of our experiments, we provide the preprocessing components of our analysis toolkit to the community at:

http://www.music.informatics.indiana.edu/code/musicxml

The archive posted at this location does not include our database (for copyright reasons), but we provide instructions for downloading the Wikifonia data set.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] M. Secrest: *Stephen Sondheim, A Life*. New York, Alfred A. Knopf, 1998.

[2] J. Peterik, D. Austin, and M. Bickford: *Songwriting for Dummies*. Hoboken, Wiley, 2002.

[3] F. Cummins: Speech Rhythm and Rhythmic Taxonomy. *Proc Speech Prosody*, April 2002.

[4] M. Brady, R. Port. Quantifying Vowel Onset Periodicity in Japanese. *Proc 16th Intl Congress of Phonetic Sciences*, Aug 2007.

[5] A.D. Patel and J.R. Daniele: An empirical comparison of rhythm in language and music.

Cognition, v87, p35-45, 2002.

[6] A.D. Patel, J.R. Iversen, and J.C. Rosenberg: Comparing the rhythm and melody of speech and music: The case of British English and French. *J. Acoustic Soc. Am*, 119(5), May 2006.

[7] S. Sapir: The intrinsic pitch of vowels: Theoretical, physiological and clinical considerations. *Journal of Voice* (3) 44-51, 1998.

[8] F. Russo, D. Vuvan, and W. Thompson: Setting words to music: Effects of phoneme on the experience of interval size. *Proc 9th Intl Conf on Music Perception and Cognition (ICMPC)*, 2006.

[9] E. Nichols: Lyric-Based Rhythm Suggestion. To appear in *Proc Intl Comp Music Conf (ICMC) 2009*.

[10] H. Oliveira, A. Cardoso, F.C. Pereira: Tra-la-Lyrics: An approach to generate text based on rhythm. *4th Intl Joint Workshop on Comp Creativity*, 2007.

[11] Y. Wang, M.-Y. Kan, T.L. Nwe, A. Shenoy, and J. Yin: LyricAlly: Automatic Synchronization of Acoustic Musical Signals and Textual Lyrics. *Proc ACM Multimedia*, Oct 2004.

[12] D. Iskandar, Y. Wang, M.-Y. Kan, H. Li: Syllabic Level Automatic Synchronization of Music Signals and Text Lyrics. *Proc ACM Multimedia*, Oct 2006.

[13] K. Lee and M. Cremer: Segmentation-Based Lyrics-Audio Alignment Using Dynamic Programming. *Proc ISMIR 2008*.

[14] F. Kleedorfer, P. Knees, and T. Pohle: Oh Oh Oh Whoah! Towards Automatic Topic Detection in Song Lyrics. *Proc ISMIR 2008*.

[15] H. Fujihara, M. Goto, and J. Ogata: Hyperlinking Lyrics: A Method for Creating Hyperlinks Between Phrases in Song Lyrics. *Proc ISMIR 2008*.

[16] R. Mayer, R. Neumayer, and A. Rauber: Rhyme and Style Features for Musical Genre Classification by Song Lyrics. *Proc ISMIR 2008*.

[17] T. Li and M. Ogihara: Music artist style identification by semi-supervised learning from both lyrics and content. *Proc ACM Multimedia*, Oct 2004.

[18] D. Wu., J.-S. Chang, C.-Y. Chi, C.-D. Chiu, R. Tsai, and J. Hsu: Music and Lyrics: Can Lyrics Improve Emotion Estimation for Music? *Proc ISMIR 2008*.

[19] Z. Eitan: *Highpoints: A Study of Melodic Peaks*. Philadelphia, Univ of Pennsylvania Press, 1997.

[20] http://speech.cs.cmu.edu/cgi-bin/cmudict Downloaded on May 20, 2009.

[21] http://dcs.gla.ac.uk/idom/ir_resources/linguistic_utils/stop_words. Retrieved on May 15, 2009.

[22] C.J. van Rijsbergen: *Information Retrieval* (2nd edition). London, Butterworths, 1979.

[23] http://en.wikipedia.org/wiki/IPA_chart_for_English_dialects. Retrieved on May 15, 2009.