"There'd be no music without the words..." — Bob Dylan (1965)

Enabling the Study of Lyric-Music Correspondence in Song: Encoding Linguistic Annotations in Humdrum Scores

Nathaniel Condit-Schultz[1]

natcs@gatech.edu

[1] *School of Music, Georgia Institute of Technology*

Motivation

Lyrics and vocals are a prominent and highly-valued element in the world's most popular music genres. Unfortunately, the Western tradition of intellectual music theory and practice has often emphasized instrumental music.

- Most computational musicology corpora focus on instrumental music.
- Many prominent corpora of vocal music omit lyrics (e.g. Essen, Bach and Prætorius chorales (Condit-Schultz, Ju, and Fujinaga 2018), Rolling Stone 200 (Temperley and Tan 2013)).
- When included, **lyric data is given less attention/detail** (e.g., poorer curation, documentation, and error checking).

"Lyrics" in ICCCM Abstracts

Counts of keywords referenced in ICCCM abstract titles and bodies.

	TOTAL N	"lyric"	"word"	"text"	"syllable"	"syntax" (language)	"harmony"	"chord"	"tonality"	"note"	"pitch"	"rhythm"	"meter"	"syntax" (music)
2023	7	0 ; 0	0 ; 0	0 ; 0	0 ; 0	0 ; 0	1 ; 3	0 ; 0	0 ; 1	0 ; 2	0 ; 2	0 ; 1	0 ; 1	0 ; 1
2024	30	0 ; 3	0 ; 2	0 ; 1	0 ; 2	0 ; 0	1 ; 4	1 ; 2	1 ; 3	0 ; 6	0 ; 4	2 ; 4	1 ; 2	0 ; 0
2025*	47	0 ; 2	0 ; 1	0 ; 2	1 ; 2	0 ; 0	4 ; 11	1 ; 10	4 ; 12	1 ; 20	5 ; 18	3 ; 10	1 ; 3	0 ; 2

* = excluding my abstract. Formatting is "N in title ; N in body".

What are "lyrics"?

- Lyrics — The **words** of a **song**, often in **written form**.
  - Composed in tandem with music or independently, and/or by same or different people (Hass, Weisberg, and Choi 2010).
- Song — Music which incorporates **human language** in the performance.
  - Language introduces abstract, **semantic meaning** into music,
    - articulated through *syntactic* structure.
  - Language also has sonic dimensions, including
    - **prosody** (rhythm/pitch, articulation) and
    - **phonetics** (pronunciation, rhyme).
    - In song, the structures of semantics/syntax, orthography, and prosody/phonetics exist in parallel to musical structures
- *Written language* introduces additional orthographic structures (punctuation, spelling, etc.).
  - As a symbolic representation, *written lyrics are the natural target of computational musicology*.
  - Huge amounts of lyric data is available online, but not aligned with melodic data.

What's in a "lyric"?

Lyrics are made up of "words," which are units of (semantic) meaning;

Written language/lyrics is focused on words, with prosodic and phonetic information underspecified or ignored.

$$\text{Orthography} \neq \text{Pronunciation} \neq \text{Prosody}$$

Western scores have adopted orthographic conventions for representing lyrics, and their relationship to musical events. However, these conventions are not consistent across time, publishers, or languages. For example, punctuation is often used ad hoc to represent musical/prosodic/syntactic units.

$$\text{Punctuation} \neq \text{Syntax} \neq \text{Prosody}$$

The most reliable prosodic information in lyric data is the syllable structure of multi-syllable words.

Consider two different encodings of a famous lyric—*can you spot the differences?*:

```
1a. I see a bad moon a ri-sin'
2a. I see trou-ble on the way

1b. I see, the Bad Moon a-ris-ing
2a. I see, troub-le on the way.
```

Lyric data often features these sorts of inconsistencies from piece-to-piece, and line-to-line.

Why do we need lyrics?

- *They are valuable in their own right.*
  - Major factor in musical taste (Demetriou et al. 2018).
  - Syllable stress affects perception of rhythm (Tan, Lustig, and Temperley 2019).
  - Influence emotional responses to music (Ali and Peynircioğlu 2006; Morton and Trehub 2007; Mori and Iwanaga 2013; Ma et al. 2024).
  - Influence musical memory (Olthof, Janssen, and Honing 2015; Racette and Peretz 2007; Peynircioğlu, Rabinovitz, and Thompson 2007; Lummis et al. 2017).

Why do we need lyrics? (continued)

- *They convey information about the rest of the music.*
  - Basic lyric features can improve performance on MIR tasks (Kim et al. 2020; Watanabe and Goto 2020), as well as more musicological tasks such as mode classification (Cornelissen, Zuidema, and Burgoyne 2020).
  - Emotional content of words are associated with musical features (Sun and Cuthbert 2017).

Music and Lyric correspondence

In the following analyses, I consider the shared information between "musical" and "lyrical" features in three datasets of sung music. This illustrates the degree to which lyric data can shed light on other aspects of the music.

Features

- *Musical features*
  - Contour (`up`, `down`, `same`)
  - Melodic interval (e.g., `±M2`)
  - Scale degree (e.g., `#4`)
  - Duration (rounded to powers of 2; e.g., `16`, `8`, `4`, `2`).
  - Procedes rest (`true` or `false`)
- *Lyrical features*
  - Position in word (`single syllable`, `first syllable`, `last syllable`, `middle syllable`)
  - Melismatic (`new syllable`, `melisma`)
  - Capitalized word
  - Punctuation (`none`, `,`, `.`)
  - Vowel (vowel characters in each syllable)

I compute "3-grams" of all features.

Datasets

- 214 songs from the *Coordinated Corpus of Popular Music* (Billboard) (Arthur and Condit-Schultz 2023)
  - Transcribed English lyrics (circa 1955–1991)
- Seven *Oratorios* by G.F. Handel, from MuseData (Selfridge-Field, Hewlett, and Sapp 2001)
  - Authorial English lyrics (circa 1708–1752)
- Ninety-one *Cantatas* by J.S. Bach, from MuseData (ibid.)
  - Authorial German lyrics (circa 1707–1750)

Mutual Information

Here, I show the mutual information (shared entropy) of each pair of features as a proportion of the joint entropy (rounded to two decimal places). This is calculated independently within each piece/movement, and the `minimum`–`mean`–`maximum` is shown.

CoCoPops

	Contour	Melodic interval	Scale degree	Duration	Procedes rest	Position in word	Melisma	Capitalized	Punctuation
Melodic interval	.54–.7–.85								
Scale degree	.28–.5–.66	.53–.7–.86							
Duration	.07–.3–.83	.16–.3–.69	.13–.3–.79						
Procedes rest	.01–.3–.62	.01–.2–.51	.01–.2–.45	.02–.3–.58					
Position in word	**.1–.2–.65**	**.13–.3–.55**	**.12–.3–.48**	**.04–.2–.57**	**.03–.2–.96**				
Melisma	**0–0–.03**	**0–0–.02**	**0–0–.02**	**0–0–.04**	**0–0–1**	.01–0–.04			
Capitalized	**.02–.1–.54**	**.02–.1–.44**	**.02–.1–.37**	**.02–.1–.47**	**.02–0–.31**	.02–.1–.34	.01–0–.27		
Punctuation	**0–0–.22**	**0–0–.23**	**0–0–.23**	**0–0–.45**	**0–0–1**	.01–0–.18	.01–.5–1	.01–.1–.27	
Vowel	**.16–.3–.48**	**.25–.4–.69**	**.22–.4–.64**	**.11–.2–.62**	**.01–.1–.27**	.12–.3–.62	0–0–.08	.02–.1–.62	0–0–.18

Handel

	Contour	Melodic interval	Scale degree	Duration	Procedes rest	Position in word	Melisma	Capitalized	Punctuation
Melodic interval	.48–.7–.91								
Scale degree	.18–.4–.77	.41–.6–.93							
Duration	.07–.2–.92	.12–.3–.78	.08–.2–.7						
Procedes rest	.05–.2–.52	.04–.2–.44	.02–.1–.4	.04–.2–1					
Position in word	**.06–.2–1**	**.11–.3–.85**	**.05–.2–.77**	**.04–.2–.92**	**.04–.1–.52**				
Melisma	**0–0–.15**	**0–0–.13**	**0–0–.12**	**.01–0–.22**	**.02–.1–.34**	.01–0–.34			
Capitalized	**.01–.1–.31**	**.01–.1–.37**	**.01–.1–.35**	**.02–.1–.4**	**.02–.1–.28**	.02–.1–.31	.02–.2–1		
Punctuation	**.04–.2–.53**	**.07–.2–.47**	**.04–.1–.39**	**.05–.2–.82**	**.07–.3–.82**	.05–.1–.4	.02–.1–.5	.02–.1–.67	
Vowel	**.1–.3–1**	**.17–.5–1**	**.1–.5–.92**	**.09–.3–.92**	**.04–.2–.52**	.12–.3–1	0–0–.15	.01–.1–.4	.06–.2–.47

Bach

	Contour	Melodic interval	Scale degree	Duration	Procedes rest	Position in word	Melisma	Capitalized	Punctuation
Melodic interval	.42–.6–.83								
Scale degree	.16–.4–.73	.37–.7–.91							
Duration	.05–.2–.47	.11–.3–.55	.08–.2–.51						
Procedes rest	.04–.2–.55	.02–.1–.38	.02–.1–.45	.04–.2–.74					
Position in word	**.05–.2–.44**	**.12–.3–.53**	**.07–.2–.52**	**.04–.1–.32**	**.03–.1–.24**				
Melisma	**0–0–.07**	**0–0–.04**	**0–0–.05**	**0–0–.12**	**.01–.1–1**	.01–0–.1			
Capitalized	**.01–.1–.29**	**.02–.1–.37**	**.01–.1–.28**	**.02–.1–.31**	**.01–.1–.26**	.03–.1–.39	.02–.1–.5		
Punctuation	**.03–.1–.27**	**.05–.2–.37**	**.04–.1–.32**	**.04–.1–.33**	**.02–.3–.92**	.06–.1–.49	.02–.1–.19	.03–.1–.28	
Vowel	**.1–.3–.63**	**.2–.5–.82**	**.09–.4–.82**	**.07–.2–.44**	**.03–.1–.27**	.2–.4–.78	.01–0–.06	.03–.2–.42	.08–.2–.63

Additional tables/analyses are shown on the `sungdrum` repo.

Encoding Recommendations and (Towards) Proposed Standards

The following can be considered *recommendations for the creation and/or curation of song datasets*—and in parallel, as *warnings of potential pitfalls for when analyzing lyric data*.

More detailed recommendations, encoding schemes, and analysis scripts are posted in my `sungdrum` repository. Currently, `sungdrum` is focused on English language, but the principles are broadly applicable.

Lyric provenance

The provenence of lyrics should be considered and encoded in metadata. Distinguish the follwing (sub)categories:

- Canonical lyrics — Associated with a *piece*
  - **Authorial**: Explicitly indicated by composer/lyricist
    - A priori (in performance score or libretto)
    - Post hoc (liner notes, Genius "verified" annotations)
  - **Documented**: Independently recorded by listeners or scholars
    - e.g., Crowd-sourced websites (pop music) or Ethnographic interviews (folk song)
- Transcribed lyrics — Associated with a *performance/recording*
  - **Aural**: By ear
  - **Mechanical**: By machine

The possibility of ambiguity or disagreement about lyrics is present in *all* forms, and must be considered.

Decoupled Encoding

It is ideal to decouple independent aspects of lyrics/sung language.

- **Distinguish orthographic information from pronunciation information.**
  - For pronunciation, International Phonetic Alphabet (IPA).
  - For word encoding, use canonical (dictionary) spellings.
    - Indicate dialect through IPA or metadata: `because` never `'cause` or `coz`.
  - Syllable boundaries should be at least consistent, and better yet systematic, but not based on pronunciation.
- **Distinguish prosodic grouping information from syntactic grouping**.
  - Omit punctuation, or use only for full stops.
  - Do not use capitalization to indicate syntactic groupings.

Humdrum Encoding Scheme(s)

The `sungdrum` repository includes detailed specifications for encoding lyric/linguistic information in humdrum-syntax data (Huron 1999).

```
**lyric         **ipa           **depends       **pos
*transcription  *LEngUS         *auto           *auto
*               *               *basic          *
*               *               !corenlp.com    !corenlp.com
=               =               =               =
i               'aI             nsub>           PRP
see             'si             root            VBP
a               @               det>moon        DT
bad             'b&d            amod>           JJ
moon            'mun            nsub>rising     NN
uh a            @               .               .
ris- -ris-      'raI            obj<see         VBG
-ing            zn              .               .
=               =               =               =
i               'aI             nsub>           PRP
see             'si             root            VBP
trou-           'trV            obj<            NN
-ble            bl              .               .
on              'an             case>way        IN
the             D@              det>            DT
way             'weI            nmod<trouble    NN
*-              *-              *-              *-
```

The repo also includes scripts for parsing and parsing humdrum lyric data, including automatic syntactic labeling via the Stanford Dependency parser (Chen and Manning 2014).

Ali, S. Omar, and Zehra F. Peynircioğlu. 2006. "Songs and Emotions: Are Lyrics and Melodies Equal Partners?" *Psychology of Music* 34 (4): 511–34. https://doi.org/10.1177/0305735606067168.

Arthur, Claire, and Nathaniel Condit-Schultz. 2023. "The Coordinated Corpus of Popular Musics (CoCoPops): A Meta-Dataset of Melodic and Harmonic Transcriptions." In *Proceedings of the International Society for Music Information Retrieval*, 239–46. https://doi.org/10.5281/zenodo.10265267.

Chen, Danqi, and Christopher Manning. 2014. "A Fast and Accurate Dependency Parser Using Neural Networks." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 740–50. Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1082.

Condit-Schultz, Nathaniel, Yaolong Ju, and Ichiro Fujinaga. 2018. "A Flexible Approach to Automated Harmonic Analysis: Multiple Annotations of Chorales by Bach and Prætorius." In *Proceedings of the International Society for Music Information Retrieval*, 66–73. https://doi.org/10.5281/zenodo.1492345.

Cornelissen, Bas, Willem Zuidema, and John Ashley Burgoyne. 2020. "Mode Classification and Natural Units in Plainchant." In *Proceedings of the International Society for Music Information Retrieval*, 869–75. https://doi.org/10.5281/zenodo.4245572.

Demetriou, Andrew, Andreas Jansson, Aparna Kumar, and Rachel Bittner. 2018. "Vocals in Music Matter: The Relevance of Vocals in the Minds of Listeners." In *Proceedings of the International Society for Music Information Retrieval*, 514–20. https://doi.org/10.5281/zenodo.1492465.

Hass, Richard W., Robert W. Weisberg, and Jimmy Choi. 2010. "Quantitative Case-Studies in Musical Composition: The Development of Creativity in Popular-Songwriting Teams." *Psychology of Music* 38 (4): 463–79. https://doi.org/10.1177/0305735609352035.

Huron, David. 1999. "Music Research Using Humdrum: A User's Guide." Stanford, California: Center for Computer Assisted Research in the Humanities.

Kim, Jaehun, Andrew M. Demetriou, Sandy Manolios, M. Stella Tavella, and Cynthia C. S. Liem. 2020. "Butter Lyrics over Hominy Grit": Comparing Audio and Psychology-Based Text Features in MIR Tasks." In *Proceedings of the International Society for Music Information Retrieval*, 861–68. https://doi.org/10.5281/zenodo.4245574.

Lummis, Sarah N., Jennifer A. McCabe, Abigail L. Sickles, Rebecca A. Byler, Sarah A. Hochberg, Sarah E. Eckart, and Corinne E. Kahler. 2017. "Lyrical Memory: Mnemonic Effects of Music for Musicians and Nonmusicians." *Psi Chi Journal of Psychological Research* 22 (2): 141–50. https://doi.org/10.24839/2325-7342.JN22.2.141.

Ma, Yiqing, David John Baker, Katherine M. Vukovics, Connor J. Davis, and Emily M. Elliott. 2024. "Lyrics and Melodies: Do Both Affect Emotions Equally? A Replication and Extension of Ali and Peynircioğlu (2006)." *Musicae Scientiae* 28 (1): 174–86. https://doi.org/10.1177/10298649221149109.

Mori, Kazuma, and Makoto Iwanaga. 2013. "Pleasure Generated by Sadness: Effect of Sad Lyrics on the Emotions Induced by Happy Music." *Psychology of Music* 42 (5): 643–52. https://doi.org/10.1177/0305735613483667.

Morton, J. Bruce, and Sandra E. Trehub. 2007. "Children's Judgements of Emotion in Song." *Psychology of Music* 35 (4): 629–39. https://doi.org/10.1177/0305735607076445.

Olthof, Merwin, Berit Janssen, and Henkjan Honing. 2015. "The Role of Absolute Pitch Memory in the Oral Transmission of Folksongs." *Empirical Musicology Review* 10 (3): 161–74. https://doi.org/10.18061/emr.v10i3.4435.

Peynircioğlu, Zehra F., Brian E. Rabinovitz, and Jennifer L. W. Thompson. 2007. "Memory and Metamemory for Songs: The Relative Effectiveness of Titles, Lyrics, and Melodies as Cues for Each Other." *Psychology of Music*, November. https://doi.org/10.1177/0305735607079722.

Racette, Amélie, and Isabelle Peretz. 2007. "Learning Lyrics: To Sing or Not to Sing?" *Memory & Cognition* 35 (2): 242–53. https://doi.org/10.3758/BF03193445.

Selfridge-Field, E., W. B. Hewlett, and C. S. Sapp. 2001. "Data Models for Virtual Distribution of Musical Scores." In *Proceedings First International Conference on WEB Delivering of Music. WEDELMUSIC 2001*, 62–70. https://doi.org/10.1109/WDM.2001.990159.

Sun, Sophia H., and Michael Scott Cuthbert. 2017. "Emotion Painting: Lyric, Affect, and Musical Relationships in a Large Lead-Sheet Corpus." *Empirical Musicology Review* 12 (3-4): 327–48. https://doi.org/10.18061/emr.v12i3-4.5889.

Tan, Ivan, Ethan Lustig, and David Temperley. 2019. "Anticipatory Syncopation in Rock: A Corpus Study." *Music Perception* 36 (4): 353–70. https://doi.org/10.1525/mp.2019.36.4.353.

Temperley, David, and Daphne Tan. 2013. "Emotional Connotations of Diatonic Modes." *Music Perception* 30 (3): 237–57. https://doi.org/10.1525/mp.2012.30.3.237.

Watanabe, Kento, and Masataka Goto. 2020. "A Chorus-Section Detection Method for Lyrics Text." In *Proceedings of the International Society for Music Information Retrieval*, 351–59. https://doi.org/10.5281/zenodo.4245442.

Computational & Cognitive Musicology Lab

Georgia Tech College of Design
School of Music