Cyber Ear / Cypher Ear, a System for Automatically Appreciating Off-the-Top Rap

Daniel Villagran
Nick Montfort
Wasalu Jaco
danielvi@mit.edu
nickm@nickm.com
wjaco@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

ABSTRACT

We describe Cyber Ear / Cypher Ear, CE/CE (pronounced "Ceecee") for short, a system for automatically evaluating improvisational, responsive rap in real time. By explicitly modeling freestyle rap appreciation, the system allows us to learn more about different aspects of this type of rap, including which ones are hard to formalize and which ones are more easily modeled computationally. Thus, this system allows us to formulate a novel theory of rap appreciation, and explore how it relies on semantic, phonetic, and lexical connections. One of the significant and difficult aspects of this type of rap is semantic connection. We compare different techniques for assessing the semantic aspects of improvisational rap and provide a preliminary assessment of these methods. We find that WordNet appears to work the best in our current implementation.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **Collaborative and social computing systems and tools**; Activity centered design; • **Computing methodologies** → **Language resources**; **Semantic networks**; • **Social and professional topics** → Cultural characteristics.

KEYWORDS

Freestyle, Rap, Cypher, Hip Hop

ACM Reference Format:
Daniel Villagran, Nick Montfort, and Wasalu Jaco. 2024. Cyber Ear / Cypher Ear, a System for Automatically Appreciating Off-the-Top Rap. In *Creativity and Cognition (C&C '24), June 23–26, 2024, Chicago, IL, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3635636.3664270

1 INTRODUCTION

Cyber Ear / Cypher Ear, CE/CE (pronounced "Ceecee") for short, is a system for automatically evaluating improvisational, responsive rap in real time as it is typed into a computer by two or more people who take turns entering four lines, or bars, at a time. By developing CE/CE, we are able to gain a better understanding of how at least one listener at a cypher might plausibly appreciate off-the-top freestyle rap lyrics. Additionally, CE/CE allows us to learn more about different aspects of this type of rap, including which ones are hard to formalize and which ones are more easily modeled computationally. Thus, this system allows us to formulate a novel theory of rap appreciation. Because semantic connection is an important part of our model, we compare different techniques for assessing the semantic aspects of improvisational rap and provide a preliminary assessment of these methods.

Coming up with bars improvisationally, or "off the top," is one of two rap practices called freestyling. Since the early days of rap in the mid-1970s, freestyling had another meaning, which involved delivering previously composed lyrics ("writtens") that were not on a single theme and did not cohere into a narrative [1]. The type of freestyle we are focusing on is that which is devised in the moment, off the top or off the dome [14]. The cypher part of CE/CE's name refers to an often spontaneous circle or session in which a group of rappers participate [4]. A circle of this sort might form at the same time each week, or might cohere in a living room or after a rap concert, outside. Even when the cypher is a regular (for instance, weekly) event, the participants will vary. A cypher may occur spontaneously, is self-governing and non-hierarchical in how it proceeds, and invites at least some improvisational rapping, even if some of what the participants deliver are writtens. Because of this, the bars in a cypher are often uniquely organic.

Our work has a similar basis to the automated player of the oral improvisational game Chain Reaction [12], which highlights differences between oral and textual traditions as theorized in [13]. While Ong's theories (about ancient societies based on orature) do not all apply to today's hyperliterate world, it is relevant to rapping in cyphers that orature is additive, aggregative, redundant, close to human lifeworld, situational, agnostically toned, and empathetic and participatory.

There have been other works on similar improvisational topics. [19] developed the bot FREESTYLE that created improvisational, responsive battle rap. Battle rap is a different, more agonistic context. This bot is also not meant to enlighten us about the appreciation of rap but rather have a machine learning algorithm that produces adequate rap, without worrying about what makes the rap adequate. [8] developed Voyager, an interactive orchestra that can respond to an improvising performer in real time, whose development is traced out in [9]. [6] studied a method called "creative arc negotiation" for

C&C '24, June 23–26, 2024, Chicago, IL, USA

ACM ISBN 979-8-4007-0485-7/24/06
https://doi.org/10.1145/3635636.3664270

creativity and improvisation. [3] conducts an "empirical study of the cognition employed by performers in improvisational theatre." [7] presents PatchProv, "a system for supporting improvisational quilt design."

We chose the cypher context so that each rapper has an opportunity to make their bars responsive to those of the previous rapper. We believe that being responsive leads to audience members finding bars "nice," that is, particularly excellent. This cypher situation also guarantees, or at least strongly incentivizes, that the rap is improvisational. If a rapper were to deliver pre-written bars when it's their turn, those would be very unlikely to respond to the previous rapper. On the other hand, actually responding to the previous rapper means that one's bars are truly improvised or at the very least, that the rapper was able to retrieve relevant written bars, which is also impressive and has elements of improvisation. In cyphers, the bars of the previous rapper give the next one something to go off of, which can make the next task easier.

2 HOW CE/CE WORKS

In the context of a cypher, CE/CE serves as an audience member. Rappers take turns typing four bars at a time into CE/CE; the systems assigns each four-bar verse a value indicating its quality. Cyphers are typically more free than this as there is not usually a set number of bars each person must rap, and people can jump in or out when they please. However, we needed a particular starting point for CE/CE. We chose four bars per person because this was a short, structured form that was used in the final round (round five) of the MC Challenge, a rapping competition held by the organization EOTW (End Of The Weak). CE/CE also gives a single grade for all four bars a rapper gives. In preliminary work, we scored individual bars. But it became clear that some bars are less responsive and/or less semantically connected because their purpose is to set up future bars that are particularly impressive and may serve as "punchlines." The "set up" bars are not particularly worse; they can be part of a verse that is very effective at a high level. In introducing a recent translation [18] of a dictionary for children by Ludwig Wittgenstein, Désirée Weber explains one of the entries that is strongly related to the English-language cypher our system engages: "*Schnaderhüpfel*, an improvised spoken-word performance that consists of at least two singers or speakers who take turns exchanging four-line stanzas. The lines are addressed from one performer to another and are often celebratory, comical, or insulting. These ... have been analogized to the Alpine version of rap battles—and are a resurgent cultural form in Austria today."

In our model, we used the following formula to quantify how a group of bars is better or worse:

$$\begin{aligned}\text{niceness} = {} & (\text{semantic association} + \text{phonetic connection})/2 \\ & - \text{lexical repetition}\end{aligned} \tag{1}$$

Phonetic connection makes lyrics musical, semantic association makes them coherent and responsive, and lexical variety keeps them fresh and varied. [14] supports this formulation because it emphasizes the importance of coherence and rhyme in rap. CE/CE uses WordNet's lemmatizer to compute lexical repetition, the CMU Pronouncing Dictionary [17] to compute phonetic repetition, and one of WordNet, Word2Vec, or BERT embeddings to compute semantic association. The final *niceness* score is a number in [0, 1].

In addition to this, we have CE/CE acknowledge that rhyme is an especially important part of rap. CE/CE will boost the score that was originally given for the presence of end rhymes in the bars of rapper. The score is given a further bonus if the rhyme is slant (or imperfect), and another further boost if the rhyme is multisyllabic.

3 RESULTS

3.1 A Theory of Rap Improvisation; CE/CE as Formalizing This Theory

Researchers have taken various theories and worked to formalize them in computational models, a process that has led to new insights. For instance, MEXICA [15] is a plot generation system that is a formal, computational model based on the Engagement-Reflection cognitive account of the creative writing process. Curveship [11] is a system for narrative variation based on several core concepts in narratology. By truly formalizing these accounts and theories, so that they become computer models, we can learn which aspects of them are rigorous and how seemingly different components may be related.

Developing CE/CE is a bit different, because there is not a pre-existing theory of off-the-top rap appreciation. We have relied, instead, on our own reflection about off-the-top rap. Fortunately, one of us is a Grammy-award-winning rapper with experience rapping off the top, and another, although much newer to rap, has been through a formal apprenticeship program and learned about rapping in this context. In this case, we have devised our theory as we have also developed a computational model. We have reason to believe that semantic connection and phonetic connection are positive aspects of rap lyrics, while avoiding lexical repetition (even of words that are inflected differently) is important.

One question we ask, then, is how hard or easy it is to detect each of these three: semantic, phonetic, and lexical connection? There are a few metrics that we can use to help give us some evidence about which computations are harder: model size, running time/computational resources needed, and development time.

- Running Time/Computational Resources: Although running time can differ based on implementation, it can offer a rough comparison between types of connection. Figure 1 displays the running time that it takes to compute each of the metrics on the Appendix A rap. Figure 2 displays the disk space that is required for models or resources that are needed to compute each of the metrics. From these results, we can infer that lexical connection is computationally the easiest, while semantic and phonetic connections are more intricate.
- Development Time: The semantics computation took the longest to develop. While lexical and phonetic repetition only have to detect multiple occurrences of the same words and phonemes respectively, there is no atomic level of semantic association whose occurrences can simply counted. Development of phonetic assessment was the second most time-consuming. While the CMU offers phonetic breakdowns of words, there are still different types of phonetic connection that have to be checked including internal rhyme, end rhyme,

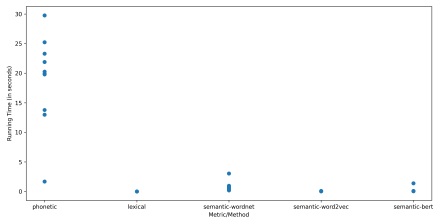

Figure 1: Running time (in seconds) to compute different aspects of connection on the verses in Appendix A. Phonetic computation is done by breaking words into phonemes and checking each phoneme against all orders, so it is $O(2^n)$. However, we might have found an alternative method; it also may not be necessary to check a window of eight bars when determining phonetic connection.

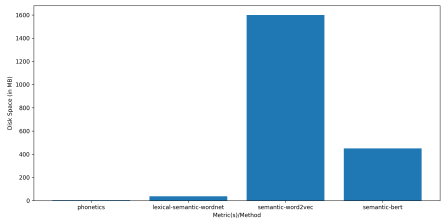

Figure 2: Disk space (in MB) to compute different aspects of connection on the verses in Appendix A. Compared to checking phonetic connection, it takes one or two orders of magnitude more data to (very approximately) determine semantic connections. Semantic and lexical connection require much more space. We use WordNet to compute lexical connection but we could have chosen to use a stemming algorithm that would have used relatively no space.

and multisyllabic rhyme. The lexical computation was the fastest and most straightforward to develop.

We understand that this is a limited view onto some aspects of complexity. There was one main developer of CE/CE, so when development time is used as a metric, we have a sample size of n=1. Nevertheless, this is an initial bit of evidence that others may be able to reinforce or refute. Our evidence suggests that ensuring semantic associations is most difficult for rappers, followed by ensuring phonetic repetition, followed by avoiding lexical repetition.

This result can be considered consistent with rappers' ideas about their practice. It is generally thought that that it is easier to make "hot garbage," or bars that sound good but make no sense. More interestingly, though, a listener is able to pick up on the sound (and some phonetic connection) even when they do not know the language. This aligns with the idea that phonetic appreciation is easier than semantic.

3.2 System Effectiveness

Due to the ephemeral nature of the improvisational, responsive rap performances we are interested in, there is no set of "canonical" works for us to use. Instead we use a performance by two of the authors, typing remotely into a shared document, on April 4, 2024, just after 2:30pm Eastern Time. Appendix A contains the text and *niceness* scores from CE/CE, using WordNet for semantic connection.

The first verse is lowest-rated, but it's hard to believe that this is the worst of the verses. CE/CE accounts for the semantic, lexical, and phonetic connections within a rapper's verse and in relation to the previous rapper's verse. For the first verse, there is no previous verse. Some sort of penalty may be reasonable here, because there is no way in our framework for this verse to be responsive.

Repeating "blue" as the final word of two bars is extremely bad ("wack," as rappers would say), but does not receive an adequate penalty because of how CE/CE currently works. The score is slightly lowered because the repetition increases the lexical repetition of the verse, but it also increases its semantic connection as any word is a synonym of itself.

The CMU pronouncing dictionary is not complete, so there are some phonetic connections that CE/CE cannot notice. For example, the first two bars in the fourth verse rhyme "OZ" with "know me." This rhyme should be significantly boosting the score of the verse as it is both slant and multisyllabic. However, the CMU pronouncing dictionary, and thus CE/CE, does not know that "OZ" should be pronounced "O-Zee."

CE/CE is also not equipped with any contextual information that is not directly mentioned in the rap. For example, the last bar of the fifth verse mentions a redeye flight being better because you can eat that time. This refers to the cypher taking place during Ramadan, and to one of the participants fasting. Even if it were possible to extract such information from text alone, we believe there are far too few hints in the text.

It's common for verses to refer to aspects of time, place, visual appearance, and even gestures made. Rappers can certainly mention current events and make references to other famous rappers; these are difficult to discern computationally even though they are purely verbal. Even more challenging, a rapper might mention the clothing that another rapper is wearing, for instance, identifying a PlayStation logo on a T-shirt and responding by rapping about how the XBox is better. It would of course be great if CE/CE could take a factor like this into account. However, this would require a change of situation — the cypher we have set up is for online participants who just type — not to mention integrating a computer vision system. We have, therefore, focused on the first practical steps toward developing an "ear" for rap.

3.3 Comparing Methods of Semantic Evaluation

CE/CE is capable of using three separate methods for computing semantic association: WordNet, Word2Vec, and BERT. We decided to choose these methods because [5], suggested WordNet, Word2Vec, and other word embeddings as effective methods for calculating semantic association.

WordNet is an online lexical database that connects words or synsets, sets of synonyms, based on semantic connections like antonym, hypernym, hyponym, etc.[2]. Word2Vec is a model that finds efficient vector representations of words [10]. BERT, Bidirectional Encoder Representations from Transformers, is a language

model that creates an embedding for a group of text (e.g. a word or a sentence) [16].

WordNet gives the first four bars of the Appendic A cypher a *niceness* score of 0.7986. However, using BERT results in a score of 0.6694 and using Word2Vec produces a score of 0.6164. Those four have quite a lot of semantic associations. Some examples are:

- fluid, moves, and lither
- big mouth, small arms, strong pecs, and T-rex
- mouth, blow, and fire

BERT should be able to pickup on semantic connections between phrases like "big mouth," "little arms," and "T-rex." However, even though BERT works beyond the word level, it, like Word2Vec, may not apply to the language in rap, which is both vernacular and poetic, because of its training data. More than that, though, since BERT compares sentences to sentences, it is not checking for semantic connections occurring within a sentence. Using to BERT to compare word-to-word instead might perform better, but needs to be studied further.

In this example, WordNet appears to score the best, but it still has its own limitations. Like Word2Vec, WordNet is not complete. For example, the last bar of the fifth verse in the cypher has the word "cuz" as a short form of "because," but WordNet does not recognize the word. In our usage of WordNet, we are checking for synonyms, antonyms, hypernyms/hyponyms, and meronyms/holonyms. We could expand this in the future to use more of WordNet's resource (e.g. checking for coordinate terms). Another issue is it can tend to pick up on connection between words that quite weak, while some more concrete connection that does not fall into a category of semantic connection that WordNet checks for. For example, in the last bar of the sixth verse, WordNet identifies "terminal" and "get" and get as being antonyms. It is taking the sense of terminal that is "at the end" and some sense of get that has to do with beginning or starting. This is quite a stretch.

4 CONCLUSION

We find that while being a promising start, there are a number of things that can be improved about our system to become a more plausible method of rap appreciation. Yet, CE/CE can still show as the semantic connection is the hardest aspect of rap to compute, which implies that this is the most difficult aspect of rap to appreciate. Lastly, we find that out of the three methods we implemented for computing semantic connection, WordNet appears to work the best for our purposes.

CE/CE is a free software, public project on GitLab. Our long-term goal for CE/CE is to make it part of a system called Cyber Rapper / Cypher Rapper (CRa/CRa), an automated improvisational, responsive rapper that would be able to participate with multiple human rappers in a cypher. Unlike CE/CE, CRa/CRa will be generator of rap, responding to the other rappers. Ideally, we could even create different, parameterized CRa/CRas, so that each one might reflect different individual and cultural backgrounds and have interesting idiosyncrasies. While the goal of this project is not to add features like obtaining voice, sound, or rhythm from audio samples to aid in the *niceness* calculation, it could be implemented by others that are interested in doing so.

ACKNOWLEDGMENTS

We thank D. Fox Harrell for discussions that led us to focus on the assessment of rap before we began work on generation.

REFERENCES

[1] P. Edwards. 2015. *The Concise Guide to Hip-Hop Music: A Fresh Look at the Art of Hip-Hop, from Old-School Beats to Freestyle Rap.* St. Martin's Publishing Group. https://books.google.com/books?id=9U1ABAAAQBAJ
[2] Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database.* Bradford Books. https://mitpress.mit.edu/9780262561167/
[3] Daniel Fuller and Brian Magerko. 2011. Shared mental models in improvisational theatre. In *Proceedings of the 8th ACM Conference on Creativity and Cognition* (Atlanta, Georgia, USA) *(C&C '11)*. Association for Computing Machinery, New York, NY, USA, 269–278. https://doi.org/10.1145/2069618.2069663
[4] Ellie M. Hisama. 2016. Improvisation in Freestyle Rap. In *The Oxford Handbook of Critical Improvisation Studies, Volume 2.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199892921.013.24 arXiv:https://academic.oup.com/book/0/chapter/295047513/chapter-ag-pdf/44511763/book_34630_section_295047513.ag.pdf
[5] Shahida Jabeen, Xiaoying Gao, and Peter Andreae. 2020. Semantic association computation: a comprehensive survey. *Artificial Intelligence Review* 53, 6 (2020), 3849–3899.
[6] Mikhail Jacob and Brian Magerko. 2021. Empirically Evaluating Creative Arc Negotiation for Improvisational Decision-making. In *Proceedings of the 13th Conference on Creativity and Cognition* (Virtual Event, Italy) *(C&C '21)*. Association for Computing Machinery, New York, NY, USA, Article 40, 10 pages. https://doi.org/10.1145/3450741.3465263
[7] Mackenzie Leake, Frances Lai, Tovi Grossman, Daniel Wigdor, and Ben Lafreniere. 2021. PatchProv: Supporting Improvisational Design Practices for Modern Quilting. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 500, 17 pages. https://doi.org/10.1145/3411764.3445601
[8] George E. Lewis. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal* 10 (12 2000), 33–39. https://doi.org/10.1162/096112100570585 arXiv:https://direct.mit.edu/lmj/article-pdf/doi/10.1162/096112100570585/1673948/096112100570585.pdf
[9] Gregor McIntosh. 2023. Man and Machine Improvising as Equals: George Lewis's Voyager. *The Elphinstone Review* (2023), 49.
[10] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781 [cs.CL]
[11] Nick Montfort. 2009. Curveship: An interactive fiction system for interactive narrating. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*. 55–62.
[12] Nick Montfort and Sebastian Bartlett Fernandez. 2022. A Minimal Computational Improviser Based on Oral Thought. In *Proceedings of the 3rd Wordplay: When Language Meets Games Workshop (Wordplay 2022)*, Marc-Alexandre Côté, Xingdi Yuan, and Prithviraj Ammanabrolu (Eds.). Association for Computational Linguistics, Seattle, United States, 16–24. https://doi.org/10.18653/v1/2022.wordplay-1.2
[13] Walter J Ong. 2002. *Orality and literacy.* Routledge, New York.
[14] Erik Pihel. 1996. A Furified Freestyle: Homer and Hip Hop. *Oral Tradition* 11 (1996), 249–269. https://api.semanticscholar.org/CorpusID:163953140
[15] Rafael Pérez y Pérez and Mike Sharples. 2001. MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence* 13, 2 (2001), 119–139. https://doi.org/10.1080/09528130010029820 arXiv:https://doi.org/10.1080/09528130010029820
[16] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/1908.10084
[17] Robert L Weide. 1998. The CMU Pronouncing Dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.
[18] Ludwig Wittgenstein and Désirée Weber. 2020. *Word Book.* Badlands Unlimited.
[19] Dekai Wu and Karteek Addanki. 2015. Freestyle: A Rap Battle Bot that Learns to Improvise. https://api.semanticscholar.org/CorpusID:17402679

A SAMPLE CYPHER WITH CE/CE'S RATINGS

Rapper A:
got some time for an online cypher
keep it fluid the moves can be lither
big mouth I blow like T-rex
small arms fire but I got strong pecs

niceness: 0.7986

Rapper B:
The force be with you young Jedi
My messages come through the air like a Redeye
If your look up you would see me when I sped by
Somewhere over the rainbow like neck tie
niceness: 0.9450

Rapper A:
I do look up to you master
notice that your plane speeds faster
and the colors shift from red to blue
tired flight can leave you blue
niceness: 0.8779

Rapper B:
I got frequent flyer miles by the OZ
That's more bounce to the ounce if you know me
I might hook you up with free flight
Or at least a couple quarters for your Street Fight
niceness: 0.8518

Rapper A:
when you travel through all the time of the night
it's not a good deal and can leave the muscles tight
but as bad as it is at least you might feel
a bit better cuz at that hour you get to have a meal
niceness: 0.9182

Rapper B:
Plane food ain't known for being the best
Some might say it can be more decent and fresh
Just get a burger at the airport in between your connect
Walk around the terminal to see what you can get
niceness: 0.8774

Rapper A:
the jetway goes to the bus of the sky
the sensation was special now you don't feel fly
the chow they serve doesn't even deserve
the name sometimes, hope the plane doesn't swerve
niceness: 0.8923

Rapper B:
Turbulence intense but its on time like Tourbillon
I use it to add some chaos into the verse that im working on
Verbal in flight service you can bet a couple purses on
Do a spree hit the duty free and get your purchase on
niceness: 0.8829

Rapper A:
you don't buy but sigh when they ask: chicken or lasagna
whatever the choice, hope it takes you into mañana
order ahead and you can get something special, a third
you had better hope it's just you, not the engine, that's intaking a bird
niceness: 0.9393

Rapper B:
The better is to bring your own snacks
Some cookies and crackers eat a whole pack
The savory sweet cashew I heard really slap
Overpriced at the Hudson News even before Tax
niceness: 0.8641

Received 8 April 2024; revised 6 May 2024