Analysing the Focus of a Hierarchical Attention Network: The Importance of Enjambments When Classifying Post-Modern Poetry

Timo Baumann[1]*, Hussein Hussein*[2]*, Burkhard Meyer-Sickendiek*[2]

[1]Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA
[2]Department of Literary Studies, Free University of Berlin, Germany

tbaumann@cs.cmu.edu, hussein@zedat.fu-berlin.de, bumesi@zedat.fu-berlin.de

Abstract

After overcoming the traditional metrics, modern and post-modern poetry developed a large variety of 'free verse prosodies' that falls along a spectrum from a more fluent to a more disfluent and choppy style. We present a method, grounded in philological analysis and theories on cognitive (dis)fluency, to analyze this 'free verse spectrum' into six classes of poetic styles as well as to differentiate three types of poems with enjambments. We use a model for automatic prosodic analysis of spoken free verse poetry which uses deep hierarchical attention networks to integrate the source text and audio and predict the assigned class. We then analyze and fine-tune the model with a particular focus on enjambments and in two ways: we drill down on classification performance by analyzing whether the model focuses on similar traits of poems as humans would, specifically, whether it internally builds a notion of enjambment. We find that our model is similarly good as humans in finding enjambments; however, when we employ the model for classifying enjambment-dominated poem types, it does not pay particular attention to those lines. Adding enjambment labels to the training only marginally improves performance, indicating that all other lines are similarly informative for the model.

Index Terms: digital humanities, free verse poetry, prosodic analysis, enjambment detection, hierarchical attention network

1. Introduction

In a groundbreaking paper on free verse poetry, Donald Wesling [1] focused on the typical line arrangements in modern poems, differing between (1) "line-sentences," as developed by Ezra Pound in *Cathay* on the basis of Ernest Fenollosa's theories of the sentence, in turn derived from the study of Chinese; (2) dismemberment of the line, whereby the line becomes "ground to the figures of its smaller units," and, as a sub-category, spatial dismemberment of the line by indentation, as William Carlos Williams does in his "triadic line verse"; (3) systematic enjambment (breaking a sentence or phrase into two lines), whereby the lines are "figures on the ground of the larger unit, the stanza"; (4) dismemberment with enjambment of the line, such that "the middle units on the rank scale engage in a protean series of identity shifts as between figure and ground" [1]. As can be seen, these classes imply a continuous development from more to less fluent styles using an increase in dismemberments and enjambments.

In this paper, we focus on the automatic detection of the enjambments. An enjambment carries over the sense or the grammatical structure from one verse line or couplet to the next without a punctuated pause. In an enjambed line (or: 'run-on-line'), the completion of a clause, phrase or sentence is delayed over to the following line so that the line ending is not emphasized as it would be in an 'end-stopped line'. In readout poetry, this delay of meaning can create a tension when the reading poet pauses before completing the sentence by reading the next line.

There are three different effects when poets read poems based on enjambments. Poets can (a) ignore the gap to the run-on-line by reading the poem fluently; (b) emphasize the gap to the run-on-line in case of a so-called "soft enjambment", which does not really affect the flow of the stanza and makes the poem still sound natural. In this case, the gap occurs between each singular colon (word group) of the poem, e.g. the noun phrase and the verbal phrase. A third option is to (c) emphasize the so-called "hard enjambments" that really interrupt the flow of the poem and its reading. This occurs when the enjambment runs across stanzas; separates articles or adjectives from their nouns or even splits a word across a line [2, 3, 4]. In modern and postmodern poetry, these techniques were developed during the so-called *free verse poetry* by modern and postmodern poets like the Imagists [5, 6, 7], and the Black Mountain poets [8, 9], who all had a strong influence on modern German poets before and after World War II [5, 10, 11, 12].

The detection of enjambment is usually based on text data, for example, the automatic detection of enjambment and its type on Spanish poems [13], in which the authors defined three kinds of enjambment: lexical (breaks up a word), phrase-bounded (phrase gets split), and cross-clause enjambment (between a relative pronoun and its antecedent). They used natural language processing (NLP) tools such as Part-of-Speech (PoS) tagger, constituency, and dependency parser. They derived 30 rules based on PoS sequences for the automatic detection of enjambment.

In this paper, we will prove that these gaps that split the enjambment from the previous line in printed verse are also audible in vocal performance. We claim that just as white spaces break up the series of black marks on the paper into smaller perceptual units whose end may or may not coincide with the end of syntactic units, oral performances may break up the text into versification units, and even indicate conflicts of versification and syntactic units.

2. Poetic Material

We collected German poems available on the webpage *lyrikline* (`www.lyrikline.org`) and the philologist and literary scholar of the project (third author), classified 175 of a total of ~2400 German poems into the six prosodic classes defined above. We also collected the corresponding audio recording of each poem as spoken by the original author, yielding a total of 52 hours of audio. Checking manually, we found some poems tagged as German that actually were not (<1 %) and discarded these from further processing.

If you look at how these poems are emphasized in the reading process, then we can identify six classes of poetic styles, within the range of fluency to disfluency: (a) The **parlando** pattern (see

Table 1: *Three examples to illustrate the free verse spectrum.*

Parlando	**Variable Foot**	**Gestic Rhythm**
Ostern am spätesten Termin,	bei geschlossenen lippen	Wie er ins Ohr dringt, dieser
an der Elbe blühte schon der Flieder,	ohne bewegung in mund und kehle	an das Herz rührende Süße enthaltende, mit ihr
dafür Anfang Dezember ein so unerhörter Schneefall,	jedes einatmen und ausatmen	erbsenschotenförmig (nur gewölbter) schwellende
dass der gesamte Bahnverkehr	mit dem satz begleiten	Ruf von da unten dem Bachbuschwerk her-
in Nord- und Mitteldeutschland	langsam und ohne stimme gedacht	
für Wochen zum Erliegen kam.	ich liebe dich	Hilferuf wohl, eines Vogels, Nestlings,

Table 2: *Description of the data used in the experiments.*

	poems	lines	characters	audio
lyrikline: German subcorpus	2392	61849	2025484	52 h
parlando	34	1435	44323	67 min
variable foot	34	878	23684	39 min
unemphasized enjambment	36	1090	33178	48 min
gestic rhythm	33	897	27741	44 min
syllabic decomposition	21	540	12390	26 min
lettristic decomposition	17	684	10007	31 min

Table 1) is based on a colon (word group)-based line arrangement. In this pattern the gap to the run-on-line, i.e. the part after the enjambment, is not emphasized in the poets reading. The second poetic pattern, (b) the **variable foot** (example in Table 1) is based on the "triadic line verse" mentioned above [2]. Like the parlando, the variable foot uses a "soft enjambment", but the poet now emphasizes the gap to the run-on-line. As long as this run-on-line also occurs between each singular colon of the poem, i.e. the noun and verbal phrases, this gap does not really affect the flow of the stanza and the poem still sounds quite natural. In the third pattern, the (c) **unemphasized enjambment**, the poet now creates a more disfluent, choppy style by using the so-called "hard enjambments" that interrupt the reading flow of the poem. This occurs when the enjambment runs across stanzas; separates articles or adjectives from their nouns or splits a word across a line. The (d) **gestic rhythm** (see Table 1) even emphasizes these hard enjambments, which makes the poem sound way more disfluent than in the two previous patterns. Even more radical kinds of poetic disfluency - below the sentence and the enjambment-level - have been developed in modern "sound poetry" by dadaistic poets like Kurt Schwitters or concrete poets like Ernst Jandl. Within the genre of sound poetry, there are two main patterns: the (e) **syllabic decomposition** and the (f) **lettristic decomposition**, the last one is the most disfluent pattern.

Some key descriptive statistics of the poems as assigned to their classes are reported in Table 2. In order to check whether poetic classes can already be singled out based on their length (in lines, characters, or audio duration) alone, we checked for significant deviations from the overall corpus. For none of the classes, the poems' durations, or number of lines significantly differ from the average poem in the corpus (two-tailed t-tests, $p > .05$ for all tests). However, variable foot poems, as well as syllabic and lettristic decompositions have significantly fewer characters than average poems.

3. Poetry Style Classification

In this section, we describe our model, which is inspired by [14], as well as our high-level decisions for modeling; more information on the model is available in [15]. Poetry, in particular post-modern poetry, is challenging material for computational modeling and statistical natural language processing. The very purpose of art (and post-modern poetry in particular) is to stand out and to defy or re-define rules, making generalization difficult. There is generally only very little data available as compared to most other domains. The automatic alignment of text and audio in spoken poetry is non-trivial (in particular in more abstract poetry) and important clues may be contained not only in how the textual material is spoken, but also in the gaps between textual material, such as extra white-space or the pausing between the lines of a poem.

Given the broad variety of the poems in combination with their relatively small number, our model must deal well with *data sparsity*. For this reason, we decide to focus our textual processing on character-by-character encoding of the lines in the poem (and using character embeddings). We use a bidirectional recurrent neural network which encodes the sequence of characters into a multi-dimensional representation that, although it is not directly applicable to human interpretation, is trained to be optimal towards differentiating the prosodic classes.

For a model to be a suitable and acceptable tool for (digital) humanistic research, it should provide insight into its decision making process, as our primary goal is not so much the automatic classification of poetry but to learn about and better understand poetic styles. To satisfy this requirement of inspectability of the decision making process (at least to some extent), we implement a notion of *inner attention* [16] as part of our encoding block that determines the final representation coming out of the decoder. Attention (a) may improve the model's representations and hence yield better performance (although some initial testing did not show a large impact), and (b) can be observed during the application of the model and gives an indication of what the model pays attention to, and can be discussed with regards to its philological plausibility. In particular, we intend to find the relationship of attention to enjambements in Section 5.

We use this encoding block (bi-RNN using gated recurrent unit (GRU) cells [17] with attention) to encode the characters of the poetic line. As for the text, we use speech line-by-line so that the model may synchronize what it 'hears' and what it 'reads'. We encode the speech (after feature extraction) using the same type of encoding block in order to capture the notion of fluency of speech delivery in the author's recitation. Finally, in order to differentiate the reading of enjambments, we also encode the pauses in-between lines, which may also contain important information about breathing, pausing, in-breath, etc. The three encodings are concatenated to form the full line representation.

We combine the line-by-line representations using a poem-level encoder (which, again, follows the same architecture) and the final representation is fed to a decision layer and a final softmax to determine the poem's class, yielding the hierarchical attention network as shown in Figure 1. While our network is

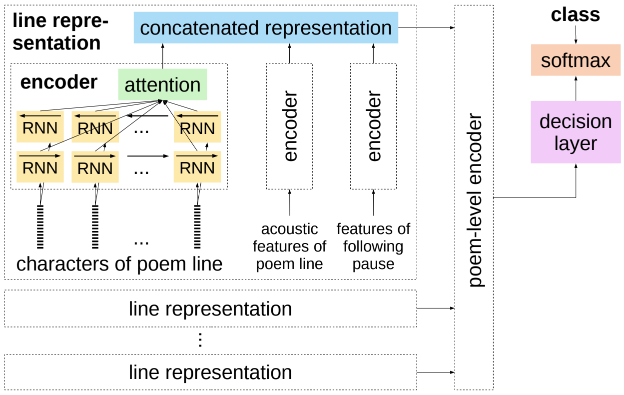

Figure 1: *Full model for poetry style detection: each line is encoded character-by-character by a RNN (using GRU cells) with attention.*

similar to those of [14] and [18], we base ours on characters instead of words as textual input and additionally include the audio stream into the analysis via additional encoders. Our model is implemented in *dyNet* [19].

3.1. Preprocessing

We perform forced-alignment of text and speech for the poems in our six classes using the text-speech aligner published by [20] which uses a variation of the SailAlign algorithm [21] implemented via Sphinx-4 [22]. The alignments are stored in a format that guarantees the original text to remain unchanged which is important to be able to recreate the exact white-spacing in the poem and would be helpful when adding further annotations (e.g. parts of speech, syntax or semantics) to the poem in the future. We extract the line-by-line timing (start of first and end of last word of the line) for each line. Forced alignment of poetry is far from trivial and often individual words cannot be aligned. Lettristic and syllabic decompositions, being a form of sound poetry, are notoriously hard to align automatically and we resorted to manual alignment of those lines that could not be aligned automatically.

We extract Mel-frequency cepstral coefficients (MFCC) for every 10 msec of the audio signal as well as fundamental frequency variation (FFV) [23] vectors, which are a continuous representation of the speaker's pitch. We z-normalize all feature dimensions. In order to not overwhelm the model with acoustic sequence information, and given that relevant speech phenomena are typically much longer than 10 msec, we compute the mean and standard deviation of 10 consecutive frames for every feature.

3.2. Pre-training

We pre-train the character embeddings and the line encoder using a recurrent autoencoder that aims to build a representation of the line that best allows it to re-create the original line (using combined costs of both forward and backward decoding as training objective); in other words: we ask our model to memorize poetic lines but given its limited memory it has to learn an abstraction of each line that helps it to remember the line.

We pre-train our representations for the acoustic features of each line similarly to the textual pre-training in that we train a recurrent autoencoder that aims to re-create the original line-by-line features, as well as the length of the acoustic stimulus. Re-creating the length of the original stimulus is particularly important as this feature is directly relevant for measuring the pause between two lines and is otherwise only a very indirect objective in pre-training. Given that line-by-line alignments are only available for the 175 poems that were manually classified, we pre-train the acoustic representations on inter-pausal units for the line representations (pausal units for between-line representations) detected using voice activity detection.

3.3. Training Procedure

Even when using pre-trained internal representations, only 175 training instances are too few for training the deep model towards the classification objective. However, poems typically display their structural properties on the vast majority of the lines they are composed of. We hence split training into two steps by first training a decision network that learns to classify individual lines of the poem in order to adapt the pre-trained network. While we here ignore the run-on-line in the case of enjambements, we do include the pausing information to model enjambments at least partially. Coming back to Figure 1, we first leave out the poem-level encoding and directly pass each line representation to a line-by-line decision layer.

Afterwards, we replace the line-by-line decision layer with the poem-level encoder and final decision layer and train towards the per-poem decisions based on the parameters estimated before. Thus, the final model is able to steer its attention mechanism towards the important lines and can learn to sacrifice the initially trained per-line optimization for the overall per-poem optimization.

For all classification experiments reported below, we perform 15 training epochs and use a dropout probability of 0.2 [24] to reduce overfitting. Each encoder is two layers deep and has a 20-dimensional state. Our character embeddings are 20-dimensional as well as the attention layers. We use 5-fold cross-validation.

4. Classifying Enjambment-dominated Poetic Styles

We use the model described in the previous section in order to (**1**) differentiate the six styles of poetic prosody, (**2**) identify poems with any of the enjambment-dominated poetic styles (variable foot, unemphasized enjambment, and gestic rhythm) from poems with any of the other styles of free-verse poetry, and (**3**) differentiate the poems' styles among the poems of enjambment-dominated styles.

The classification results in terms of f-measure are presented in Table 3. We find **1:** that classifying post-modern free-verse poems into their prosodic classes is feasible with our model, yielding an f-measure of 0.73 for the 6-class task. Furthermore, **2:** we notice that the three enjambment-related classes can be differentiated from the other three classes perfectly and that **3:** the 3-class task of differentiating the enjambment-dominated poems yields relatively low performance. One would expect, in contrast, better performance on the 3-class problem over the 6 classes. This indicates that the enjambment-dominated classes are harder to differentiate from each other than the other three. We next investigate how the model's performance relates to its notion of enjambment and whether an explicit notion of enjambment helps to classify enjambment-dominated styles.

Table 3: *Results (weighted f-measure) for classification tasks*

	classification task	f-measure	relative attention on enjambments
1.	classify into 6 poetic-prosodic styles	0.73	
2.	identify enjambment-dominated poems	1.	
3.	classify 3 styles of enjambment-dominated poems	0.69	0.98
4.	classify enjambment lines including ‘unclear’ cases	0.69	—
5.	classify enjambment lines excluding ‘unclear’ cases	0.91	—
6.	classify 3 styles with explicit notion of enjambmnt	0.70	1.

5. Investigating Enjambments

Above, we have dealt with the identification, as well as the differentiation of poems that are dominated by enjambments. In this section, we take a look at the importance of the individual enjambments, i. e. at those lines in a poem which run over in meaning and/or syntax to the next.

For the 103 poems in our corpus that are dominated by enjambment, we had two annotators (an expert in literary studies and a literarily unbiased native speaker of German not involved in this research) annotate for each line whether it was an enjambment or not (or whether they faced a severe difficulty in taking this decision).

Of the total of 2286 annotated lines, 435 lines were labeled as unclear by either annotator (most by the expert) and are excluded from the analysis below. We measure the inter-annotator agreement for those cases that are more clear and find a Cohen's kappa of $\kappa = 0.89 \pm 0.01$ and differences in less than 6 % of the lines. Overall, some 59 % of the lines have been annotated as enjambments. We use these line-by-line anotations in two ways, as described in the next paragraphs.

Based on the hypothesis – motivated by literary study – that the enjambment-dominated classes differ in their use of enjambment, we speculate that the classifiers' attention mechanism would pay particular attention to enjambment lines. We test this hypothesis by recording the attention of the model on every line (normalizing for the length of the poem) and comparing the resulting attentions to whether a line is an enjambment or not. We call the average of these values the *relative attention* of the model on enjambment/end-stopped lines. We expect that the relative attention should be larger than 1 for enjambment lines (as these are deemed to be critical to the decision making process by the literature) and lower for end-stopped lines. In contrast, if the model distributes its attention similarly across all lines, we expect relative attention to be close to 1. As shown in Table 3 (right-most column), relative attention is 0.98 and hence not focused on enjambment lines but spread out almost equally across the three line annotations (enjambment, no enjambment, unclear). We thus conclude that the model performs classification *without focusing* on enjambment lines, unlike postulated by literary study.

There are several reasons why our model may not focus on enjambment lines for differentiating enjambment poems: (a) it may be unable to find enjambment lines, (b) the underlying recurrent neural network may well encode the occurrence of an enjambment line and carry this to a next state which is then favored by the attention mechanism, (c) enjambment lines do not actually differentiate the styles in question.

While ruling out (b) would require a deep analysis of the network's parameters, we can test (a) easily by building a classifier that differentiates enjambments from lines that are not. We first build a classifier that differentiates the lines into the three classes ‘clear enjambment’, ‘clear non-enjambment’ (where both annotators agree), and ‘unclear case’ (where annotators disagree or explicitly mark this an unclear case). **4:** The resulting weighted f-measure is .69. In particular, the f-measure for the ‘unclear’ class is low at .42. We hence also build a binary classifier to differentiates only the clear cases and in this case reach **5:** an f-measure of 0.91, mirroring human annotation performance. Thus, we conclude that the neural network is well *able* to differentiate between enjambements and non-enjambments. It simply did not choose to do so in the end-to-end training.

We finally include explicitly whether a line is an enjambment as a binary feature into the concatenated representation of each line (the blue box in Figure 1). We then retrain another classifier to differentiate the 3 classes. We find **6:** a marginal improvement of the performance and no particular focus of the model on those lines (relative attention is exactly 1).

6. Conclusion and Future Work

We have investigated the influence of individual enjambment lines when classifying different enjambment-dominated poetic styles in post-modern free-verse poetry. Among our classification hierarchy of 6 poetic styles for postmodern free-verse poetry, we found enjambment-dominated styles particularly difficult to differentiate. We hence tested whether humans can reliably annotate enjambments (they can) and whether our model is able to identify such lines (it can).

Yet, we found – as evidenced by the model's attention behaviour – that the model neither builds an implicit notion of enjambment nor makes use of the explicit notion of enjambment if it is provided while classifying the enjambment-dominated styles. Explicit information about enjambments only leads to a marginal improvement in classification performance. This could be for several reasons: either the model encodes enjambments slightly differently such that the attention mechanism does not allow us to recover its internal state (for example, we are measuring the attention of the line that is interrupted but conceivably the model might attend to lines following the enjambment). Alternatively, we must reconsider our philological notion that the styles in question are primarily differentiated by the characteristics of the enjamb'ed lines.

7. Acknowledgements

This work is funded by Volkswagen Foundation in the announcement ‘Mixed Methods in the Humanities.’ We thank Verena Willkomm for providing a line-by-line annotation of the enjambments.

8. References

[1] D. Wesling, "The Prosodies of Free Verse," in *Twentieth-Century Literature in Retrospect*, R. A. Brower, Ed. Cambridge, MA: Harvard University Press, 1971, pp. 155–187.

[2] S. Cushman, *William Carlos Williams and the Meaning of Measure*, ser. Yale Studies in English. New Haven and London, 1985.

[3] C. Wagenknecht, *Deutsche Metrik: eine historische Einführung*. Beck, 1999.

[4] R. Greene, S. Cushman, C. Cavanagh, J. Ramazani, P. Rouzer, H. Feinsod, D. Marno, and A. Slessarev, *The Princeton Encyclopedia of Poetry and Poetics*, ser. Princeton reference. Princeton University Press, 2012.

[5] J. Silkin, *The Life of Metrical and Free Verse in Twentieth-Century Poetry*. Palgrave Macmillan UK, 1997.

[6] G. B. Cooper, *Mysterious Music: Rhythm and Free Verse*. Stanford University Press, 1998.

[7] C. Beyers, *A History of Free Verse*. University of Arkansas Press, 2001.

[8] A. Golding, "Charles Olson's Metrical Thicket: Toward a Theory of Free-Verse Prosody," *Language and Style*, vol. 14, pp. 64–78, 1981.

[9] T. Steele, *Missing Measures: Modern Poetry and the Revolt Against Meter*. University of Arkansas Press, 1990.

[10] A. Finch, *The Ghost of Meter: Culture and Prosody in American Free Verse*. University of Michigan Press, 2000.

[11] B. Meyer-Sickendiek, *Lyrisches Gespür: Vom geheimen Sensorium moderner Poesie*. Fink, Wilhelm, 2012.

[12] J. Lüdtke, B. Meyer-Sickendiek, and A. M. Jacobs, "Immersing in the stillness of an early morning: Test ing the mood empathy hypothesis of poetry reception," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 8, no. 3, pp. 363–377, 2014.

[13] P. Ruiz Fabo, C. Martínez Cantón, T. Poibeau, and E. González-Blanco, "Enjambment Detection in a Large Diachronic Corpus of Spanish Sonnets," in *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Association for Computational Linguistics, August 2017, pp. 27–32.

[14] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480–1489.

[15] T. Baumann, H. Hussein, and B. Meyer-Sickendiek, "Style Detection for Free Verse Poetry from Text and Speech," in *Proceedings of COLING 2018*, August 2018.

[16] Y. Liu, C. Sun, L. Lin, and X. Wang, "Learning natural language inference using bidirectional LSTM model and inner-attention," *CoRR*, vol. abs/1605.09090, 2016. [Online]. Available: http://arxiv.org/abs/1605.09090

[17] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1724–1734. [Online]. Available: http://www.aclweb.org/anthology/D14-1179

[18] A. Tsaptsinos, "Lyrics-based music genre classification using a hierarchical attention network," in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, 2017, pp. 694–701.

[19] G. Neubig, C. Dyer, Y. Goldberg, A. Matthews, W. Ammar, A. Anastasopoulos, M. Ballesteros, D. Chiang, D. Clothiaux, T. Cohn, K. Duh, M. Faruqui, C. Gan, D. Garrette, Y. Ji, L. Kong, A. Kuncoro, G. Kumar, C. Malaviya, P. Michel, Y. Oda, M. Richardson, N. Saphra, S. Swayamdipta, and P. Yin, "Dynet: The dynamic neural network toolkit," *arXiv preprint arXiv:1701.03980*, 2017.

[20] T. Baumann, A. Köhn, and F. Hennig, "The Spoken Wikipedia Corpus Collection: Harvesting, Alignment and an Application to Hyperlistening," *Language Resources and Evaluation*, 2018.

[21] A. Katsamanis, M. Black, P. G. Georgiou, L. Goldstein, and S. Narayanan, "SailAlign: Robust long speech-text alignment," in *Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*, 2011.

[22] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, "Sphinx-4: A Flexible Open Source Framework for Speech Recognition," Mountain View, CA, USA, Tech. Rep., November 2004.

[23] K. Laskowski, M. Heldner, and J. Edlund, "The fundamental frequency variation spectrum," in *Proceedings of FONETIK 2008*, 2008.

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.