Analysis of Rhythmic Phrasing: Feature Engineering vs. Representation Learning for Classifying Readout Poetry

Timo Baumann

Carnegie Mellon University
Language Technologies Institute

Hussein Hussein
Burkhard-Meyer-Sickendiek

Freie Universität Berlin

Analysis of Rhythmic Phrasing:
Feature Engineering vs. Representation Learning
for Classifying Readout Poetry

in particular: (post)modern free-verse poetry

Rhythmicalizer project

- **Objective:** Automatic classification of rhythmical patterns in readout poetry based on a „free verse prosody“

- **Project:**
  Duration: 2017-2020
  Funding by the Volkswagenstiftung in the program „Mixed Methods in the Humanities“

- **Cooperation Partner:**
  Literaturbrücke e.V. with its webpage lyrikline.org

Lyrikline: author-spoken post-modern poetry

> 11000 poems

> 1200 poets

> 80 languages (primarily DE & EN)

- many poetic translations

The fluency spectrum of free verse poetry

- **Fluency** of cognitive processing causes positive effect of aesthetic experience (Topolinski and Strack, 2009)
- Stimuli that are processed more easily get higher rankings (Belke et al., 2010)

- Abstract art (Picasso or Schönberg) created disfluency as an artistic strategy (Bullot and Reber, 2013)
- **Disfluency** prompts people to process information on a higher level of abstraction (Smith and Smith, 2006)

The fluency spectrum of free verse poetry

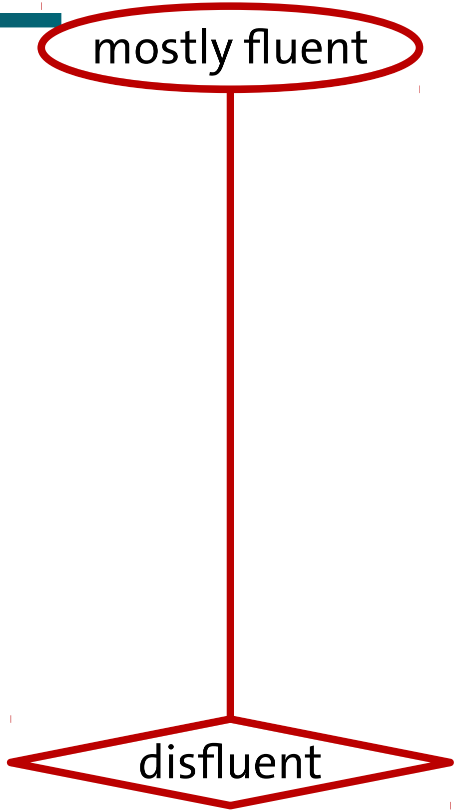
- **Fluency** of cognitive processing causes positive effect of aesthetic experience (Topolinski and Strack, 2009)
- Stimuli that are processed more easily get higher rankings (Belke et al., 2010)

- Abstract art (Picasso or Schönberg) created disfluency as an artistic strategy (Bullot and Reber, 2013)
- **Disfluency** prompts people to process information on a higher level of abstraction (Smith and Smith, 2006)

The fluency spectrum of free verse poetry

literary stylistic classes along the continuum:

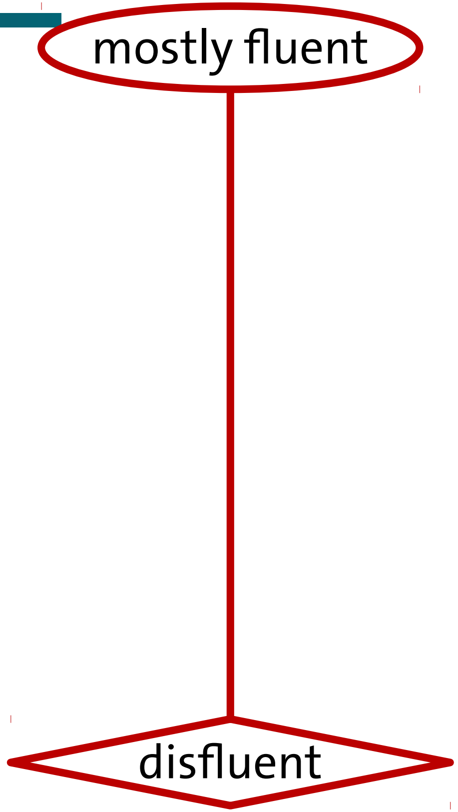
- Parlando style
- Variable foot
- Unemphasized enjambment
- Gestic rhythm
- Syllabic decomposition
- Lettristic decomposition



The fluency spectrum of free verse poetry

literary stylistic classes along the continuum:

- Parlando style
- Variable foot
- Unemphasized enjambment
- Gestic rhythm
- Syllabic decomposition
- Lettristic decomposition



Audio examples

literary stylistic classes along the continuum:

• Parlando style	Benn: TEILS-TEILS
• Variable foot	Jandl: beschreibung eines gedichts
• Unemph. enjamb.	Kirsten: BEINHART
• Gestic rhythm	Hensel: ALS ICH BEI IHM WAR
• Syllabic dec.	Jandl: auf dem lande
• Lettristic dec.	Jandl: schtzngrmm

Stylistic classes along the spectrum

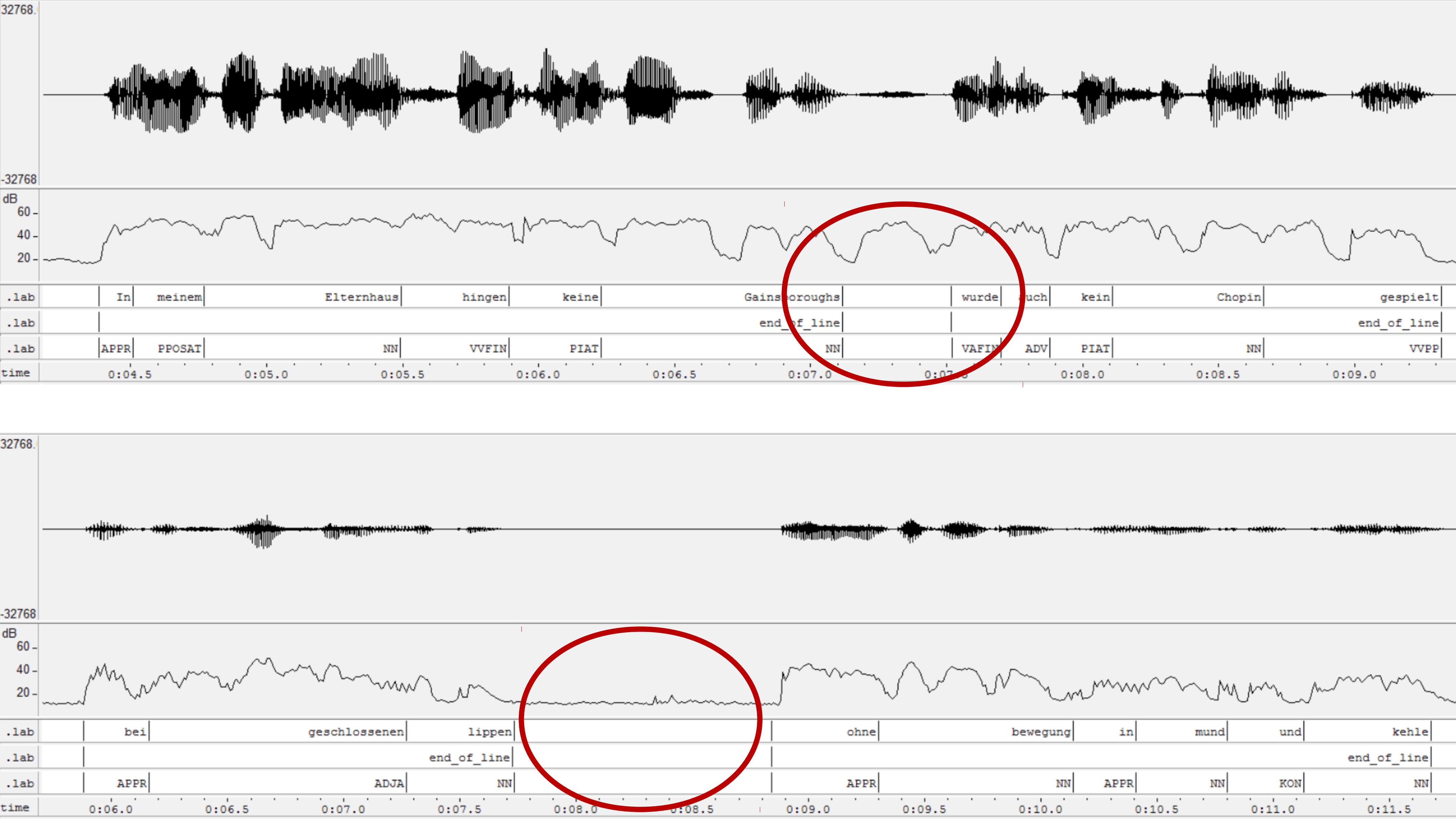
- literature discusses styles of poetry
  - theories about relations of styles, e.g.:
- Rhythmic Phrasing in English Verse (Cureton 1992)
  - based on theory in tonal music
  - rhythm consists of
    meter: perception of beats in regular patterns
    grouping: linguistic units around a single climax
    prolongation: anticipation and overshooting of a goal

Stylistic classes along the spectrum

- literature discusses styles of poetry
  - theories about relations of styles, e.g.:
- Rhythmic Phrasing in English Verse (Cureton 1992)
  - based on theory in tonal music
  - rhythm consists of
    meter: perception of beats in regular patterns
    grouping: linguistic units around a single climax
    prolongation: anticipation and overshooting of a goal

Parlando and Variable Foot

- **similar grouping** in lines of phrases/clauses
  - frequent use of enjambments,
    breaking of contextual units across lines
- **different prolongation**:
  - Parlando flows smoothly across line-breaks
  - Variable foot uses breath-controlled lines and
    pauses between lines
- Parlando uses “more fluent” **spoken form**


32768
-32768
dB
60
40
20
.lab
In
meinem
Elternhaus
hingen
keine
Gains oroughs
wurde
uch
kein
Chopin
gespielt
end of_line
end_of_line
APPR
PPOSAT
NN
VVFIN
PIAT
NN
VAFIN
ADV
PIAT
NN
VVPP
time
0:04.5
0:05.0
0:05.5
0:06.0
0:06.5
0:07.0
0:08.0
0:08.5
0:09.0
32768
-32768
dB
60
40
20
.lab
bei
geschlossenen
lippen
ohne
bewegung
in
mund
und
kehle
end_of_line
end_of_line
APPR
ADJA
NN
APPR
NN
APPR
NN
KON
NN
time
0:06.0
0:06.5
0:07.0
0:07.5
0:08.0
0:09.0
0:09.5
0:10.0
0:10.5
0:11.0
0:11.5

Experiment:

- classify style based on text + audio
  - features that (partially) align to theory
  - theory-agnostic representation learning with Nns
- forced alignment of text + speech
  POS tagging with Stanford parser (STTS tagset)

Data

- German poems
- 68 poems (34 each per class Parlando and VarF)
- 24 authors (reading 1-9 poems each)
- 9-200 lines / 27-700 seconds per poem

Feature-Engineering and Classification

- motivated by theory:
  - avg. pause duration between lines
    → captures strength of enjambments
  - avg. pausing between words
    → indirectly normalizes for speech tempo
- additionally:
  - there may be differences in phrasing
    → #lines, #with finite verb, #with punctuation

Results for feature-based Classification:

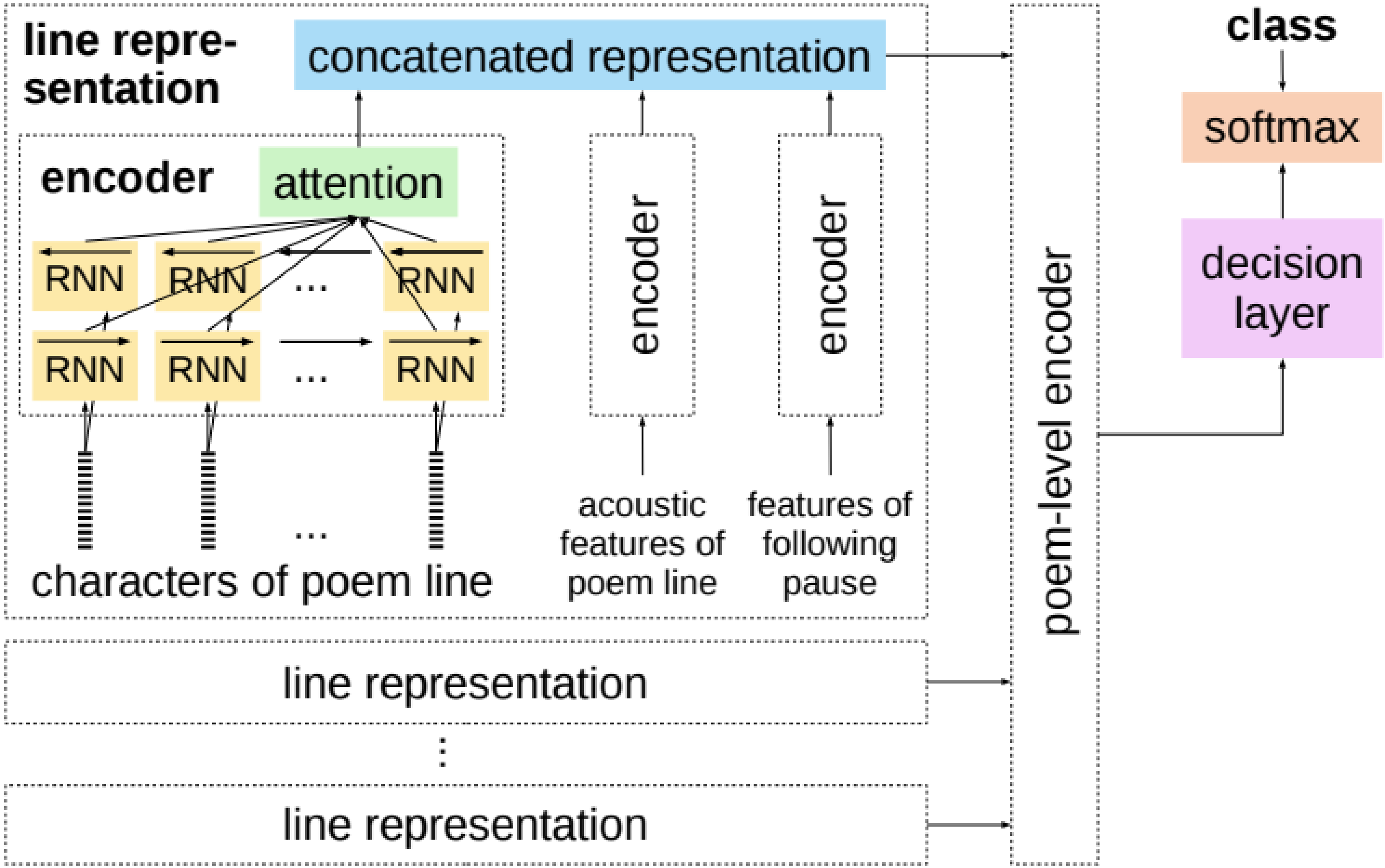
Table 1: *Experimental results (weighted average of f-measure) obtained with the 10 fold cross-validation by applying different feature sets on several classification algorithms.*

Classifier \ Features	Pause	Parser	Pause & Parser
AdaBoostM1	0.59	**0.69**	0.62
IBk	**0.68**	0.68	0.59
SimpleLogistic	0.47	0.63	**0.66**
RandomTree	0.65	0.56	0.53

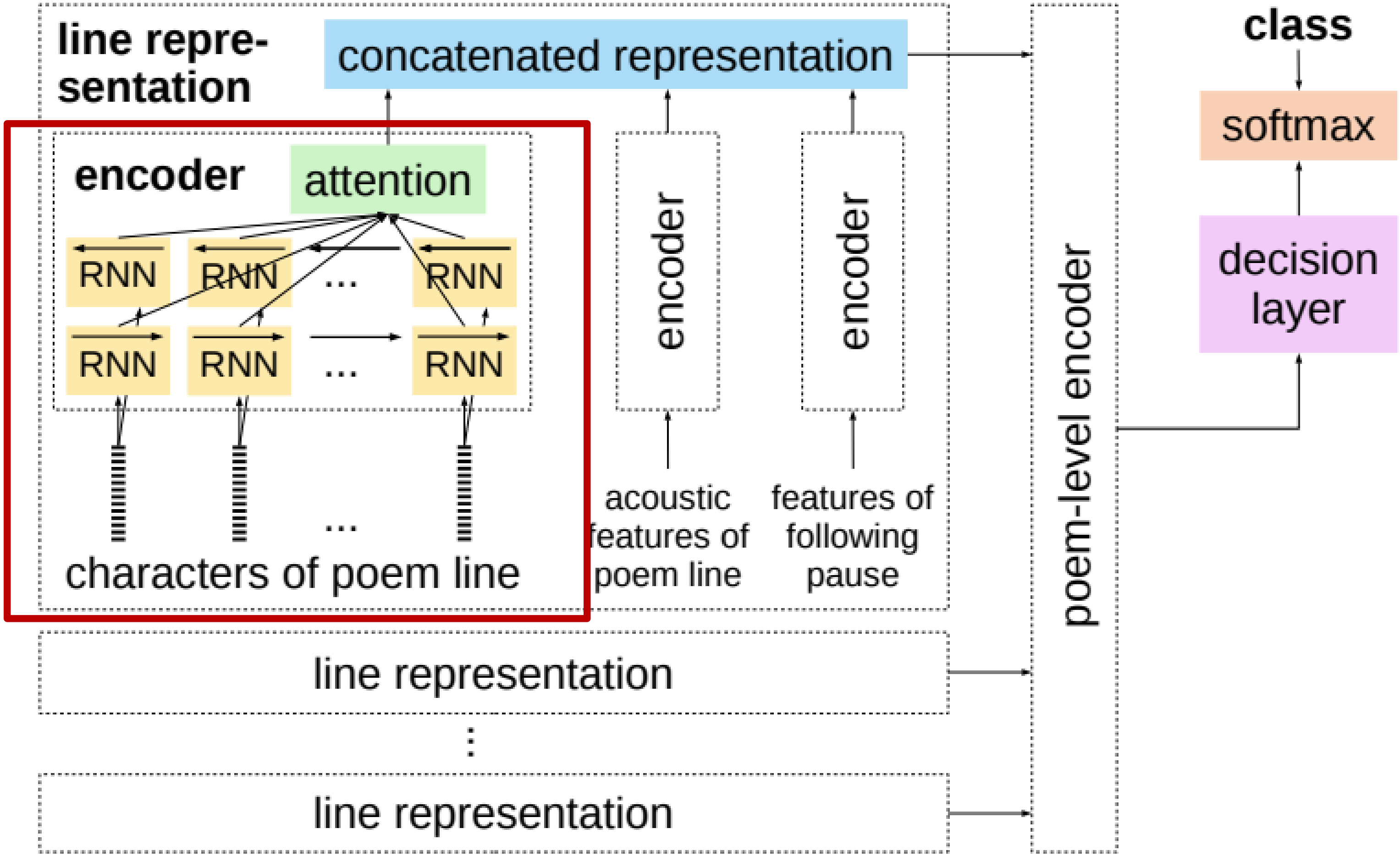
Analysis

- “better than chance” classification
- large variation between diff. classifiers
  - probably indicative of instability of results
- very little data → just 68 poems
- few features
  (but what would be better features?)

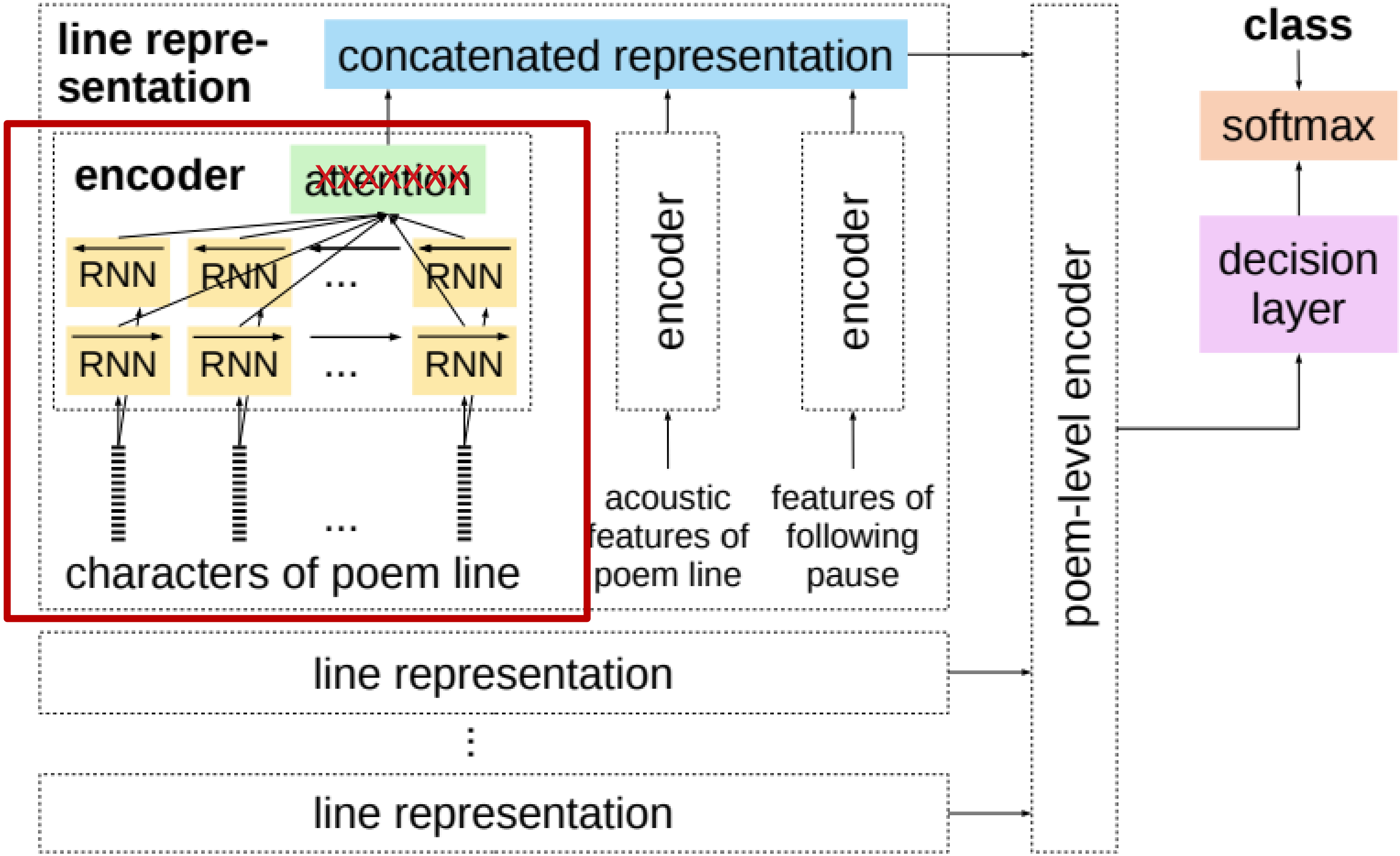
Hierarchical and Recurrent Neural Network



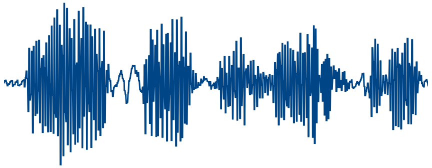
Hierarchical and Recurrent Neural Network



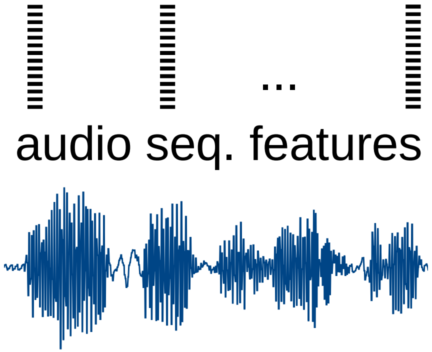
Hierarchical and Recurrent Neural Network



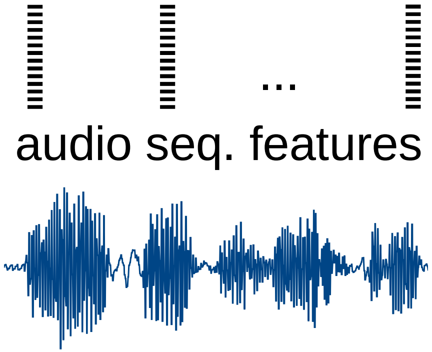
RNN modeling of a sequence



RNN modeling of a sequence

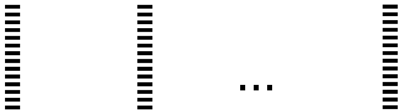


RNN modeling of a sequence

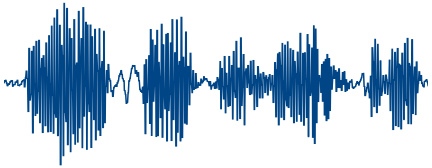


MFCC + FFV

RNN modeling of a sequence



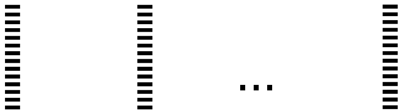
audio seq. features



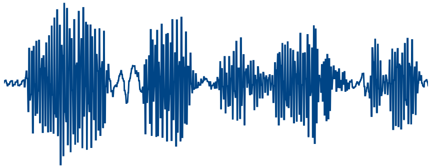
MFCC + FFV

standard speech features

RNN modeling of a sequence



audio seq. features

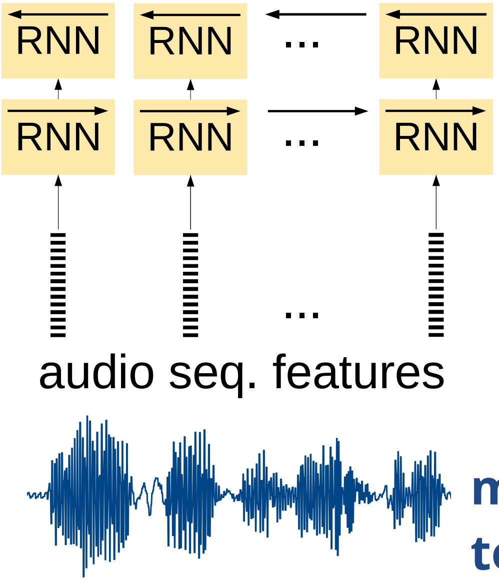


MFCC + FFV

standard speech features	**character embeddings for text**

model can relate prosody (or text) to poetic style

RNN modeling of a sequence



bi-directional LSTMs work better than uni-directional LSTMs.

MFCC + FFV

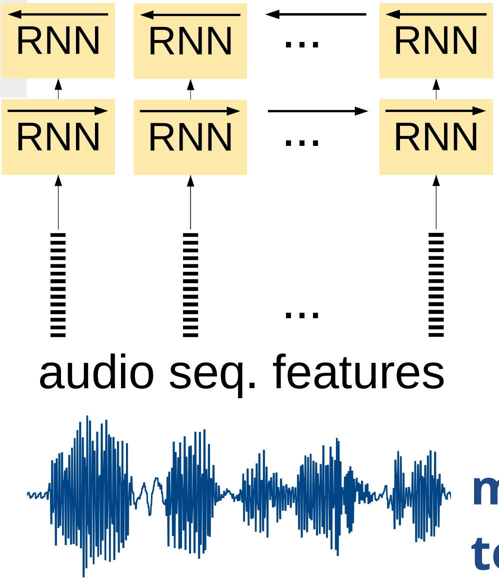
standard speech features

character embeddings for text

model can relate prosody (or text) to poetic style

RNN modeling of a sequence

gets fixed-length representation from variable-length sequence



bi-directional LSTMs work better than uni-directional LSTMs.

MFCC + FFV

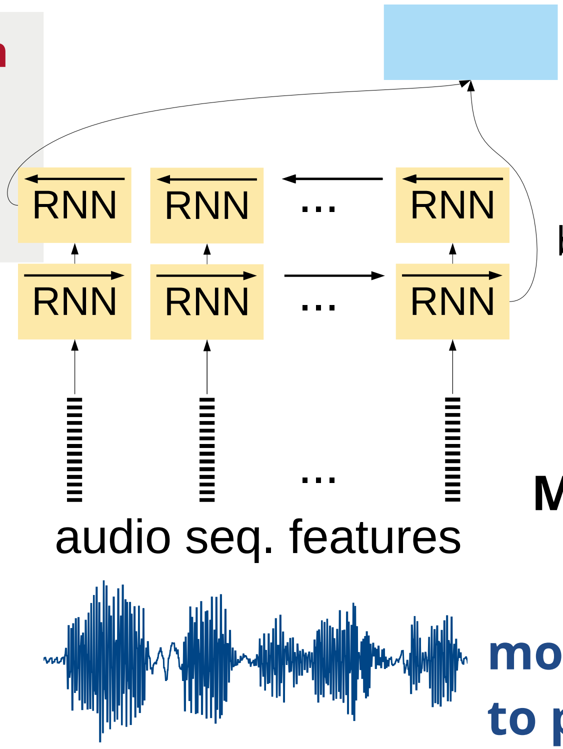
standard speech features

character embeddings for text

model can relate prosody (or text) to poetic style

RNN modeling of a sequence

gets fixed-length representation from variable-length sequence



a multi-dimensional representation of the poetic line

bi-directional LSTMs work better than uni-directional LSTMs.

MFCC + FFV

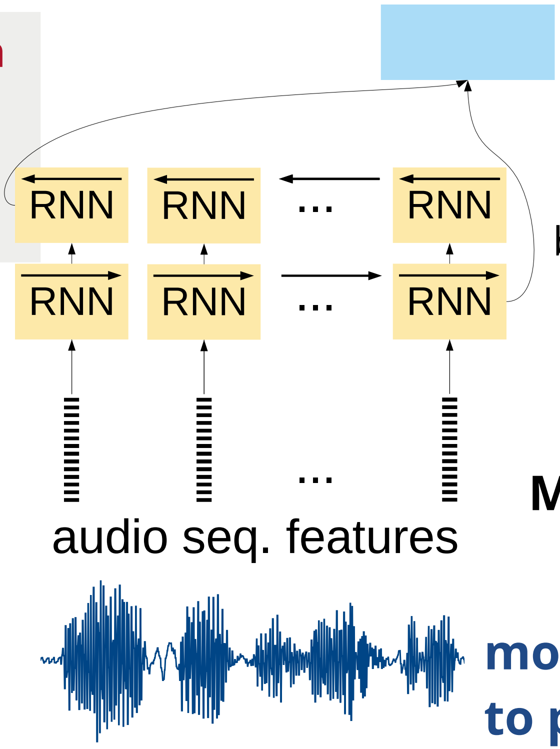
standard speech features

character embeddings for text

model can relate prosody (or text) to poetic style

RNN modeling of a sequence

gets fixed-length representation from variable-length sequence



a multi-dimensional representation of the poetic line

unfortunately not easily interpretable

bi-directional LSTMs work better than uni-directional LSTMs.

MFCC + FFV

standard speech features

character embeddings for text

model can relate prosody (or text) to poetic style

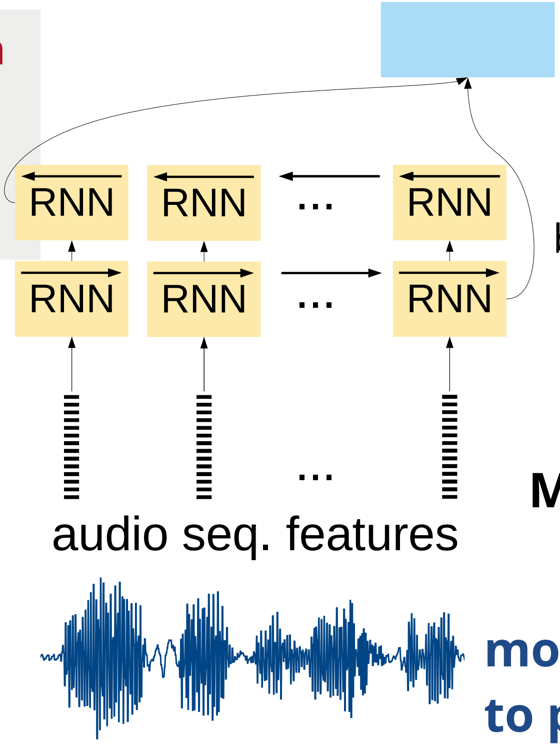
RNN modeling of a sequence

class label

a multi-dimensional representation of the poetic line

unfortunately not easily interpretable

gets fixed-length representation from variable-length sequence



bi-directional LSTMs work better than uni-directional LSTMs.

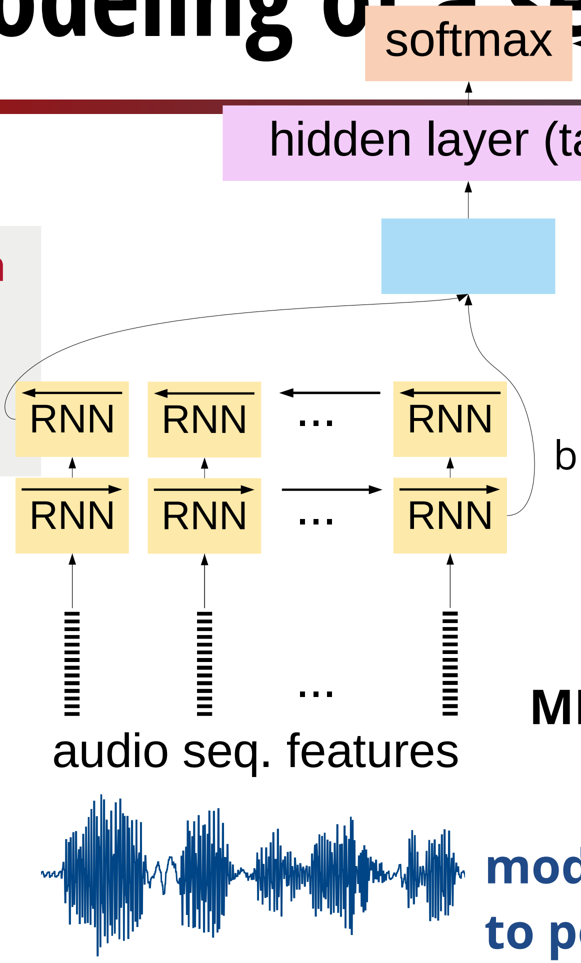
MFCC + FFV

standard speech features

character embeddings for text

model can relate prosody (or text) to poetic style

RNN modeling of a sequence



gets fixed-length representation from variable-length sequence

a multi-dimensional representation of the poetic line

unfortunately not easily interpretable

bi-directional LSTMs work better than uni-directional LSTMs.

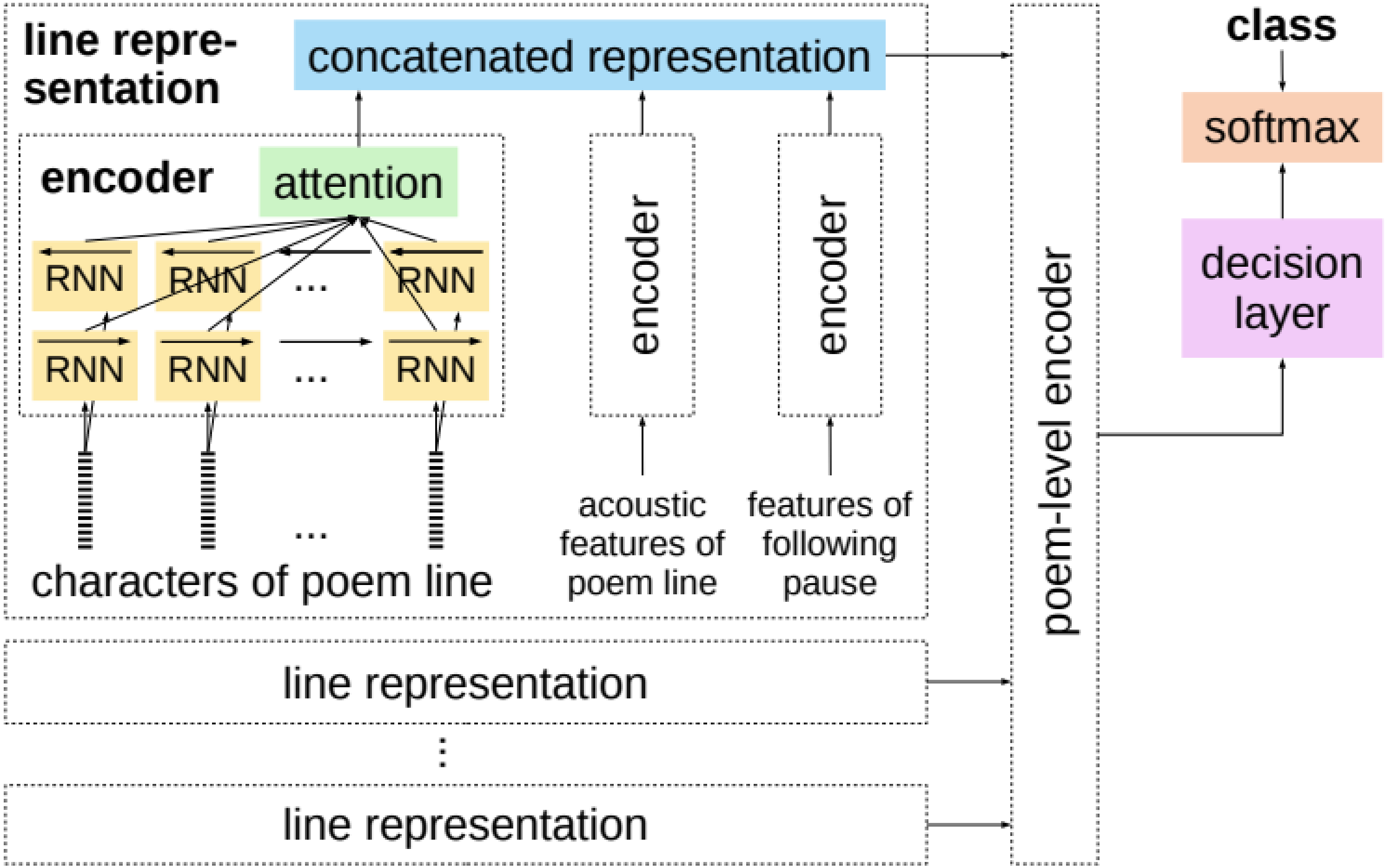
MFCC + FFV

standard speech features

character embeddings for text

model can relate prosody (or text) to poetic style

Hierarchical and Recurrent Neural Network



Neural networks-based approach

- NNs are good at identifying hidden patterns
- but typically require large amounts of data
  - only 68 poems!!
  - but together they consist of 2313 lines
- two-stage training:
  - pretrain a model that classifies lines
  - use pretraining results to train poem classification

Training procedure for Neural Network:

- based on the forced-alignment, use:
  - characters of the line
  - audio of the line
  - audio of the following pause

  to train a line-by-line classifier
  - results in reasonably pre-trained line-representations
- train another line-to-poem encoder for full poem classification

Training setup:

- keep #parameters low
  - low dimensional character encodings
  - GRUs (fewer parameters than LSTMs)
- dropout to avoid overfitting

- 10-fold crossvalidation

Results and Comparison

Table 1: Results (weighted f-measure) for both approaches.

classifier and feature engineering			NN and representation learning		
pause	parser	pause+parser	text	text+speech	text+speech+pause
0.59	0.69	0.62	0.65	0.85	0.85

- Hierarchical network outperforms
- most value seems to be in the speech
- prolongation not only expressed by pause

Extension of the NN to 6 classes (Coling paper)

	f-measure	parlando	var. foot	unemph. enj.	gestic	syll. dec.	lettr. dec.
parlando	0.83	30	2	2			
variable foot	0.60	3	20	6	5		
unemph. enj.	0.71	2	4	27	3		
gestic rhythm	0.68		6	5	21		1
syllabic dec.	0.81	2	1			17	1
lettristic dec.	0.77	1				4	12

	all features	no pause	text-only
all six classes	0.73	0.66	0.47
parlando vs. variable foot	0.85	0.85	0.65
unemphasized enjambment vs. gestic rhythm	0.78	0.66	0.57
syllabic vs. lettristic dec.	0.82	0.92	0.82

Rhythmicalizer project

Lyrik-
—line listen to the poet

Thank you!

VolkswagenStiftung

Carnegie Mellon University
Language Technologies Institute

www.timobaumann.de/work

Rhythmicalizer project

Thank you!

Thank you!

Lyrik-
line
listen to the poet

Thank you!

Thank you!

Thank you!

VolkswagenStiftung

Carnegie Mellon University
Language Technologies Institute

www.timobaumann.de/work

Character-based embeddings?

Character-based embeddings?

- can character-based embeddings be used to attest high-level phenomena?

  → the RNN step recombines the characters to words (or other more abstract concepts)
- still, why not use word-embeddings?

  →consider decomposition poetry:

Character-based embeddings?

- can character-based embeddings be used to attest high-level phenomena?

  → the RNN step recombines the characters to words (or other more abstract concepts)

- still, why not use word-embeddings?

  →consider decomposition poetry:

rininininininininDER
brüllüllüllüllüllüllüllüllEN
schweineineineineineineineinE
grunununununununununZEN

```
der minister bedauert d r rt g    ss r ng n
der minister bedau rt d r rt g    ss r ng n
der minister bed   rt d r rt g    ss r ng n
```

schtzngrmm
schtzngrmm
t-t-t-t
t-t-t-t

Timo Baumann