On the Importance of Time-dependent Features for Speaker Recognition

Master Thesis

Daniel Neururer

Zurich University of Applied Sciences

October 20, 2020

Table of Contents

Introduction

Speaker Recognition

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Modelling a representation of a speaker

- Speaker Identification
- Speaker Verification
- Speaker Clustering

- What needs to be considered?

Introduction

Speaker Recognition

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Modelling a representation of a speaker

- Speaker Identification
- Speaker Verification
- Speaker Clustering

- What needs to be considered?

Introduction

Speaker Recognition

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Modelling a representation of a speaker

- Speaker Identification
- Speaker Verification
- Speaker Clustering

- What needs to be considered?

Introduction

Attributes of Speech

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Acoustic Features
  - Pitch
  - Loudness
  - Timbre
- Prosodic Features
  - Intonation
  - Stress
  - Speech Tempo
  - ...

Introduction

Attributes of Speech

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Acoustic Features
  - Pitch
  - Loudness
  - Timbre
- Prosodic Features
  - Intonation
  - Stress
  - Speech Tempo
  - ...

Introduction

Relevance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Most valuable Information encoded in Acoustic Features

- DNN's might focus too much on learning acoustic features
- DNN's might ignore temporal information

- Experiment Time!

Introduction

Relevance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Most valuable Information encoded in Acoustic Features

- DNN's might focus too much on learning acoustic features
- DNN's might ignore temporal information

- Experiment Time!

Introduction

Relevance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Most valuable Information encoded in Acoustic Features

- DNN's might focus too much on learning acoustic features
- DNN's might ignore temporal information

- Experiment Time!

Introduction

Relevance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Most valuable Information encoded in Acoustic Features
- DNN's might focus too much on learning acoustic features
- DNN's might ignore temporal information
- Experiment Time!
- 70% of the information is a gift

Methods

DNN Architectures

School of Engineering
InIT Institut für angewandte Informationstechnologie

- LUVO
- LSTM
- ResNet34S

- Time Independent ResNet34S
- MLP

Methods

DNN Architectures

School of Engineering
InIT Institut für angewandte Informationstechnologie

- LUVO
- LSTM
- ResNet34S

- Time Independent ResNet34S
- MLP

Methods

Evaluation Methods

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Use one Segment only
- Use consecutive Segments with a Hop Percentage
  Calculate the Arithmetic Mean of the Embeddings
- Use the full Utterance

Methods

Evaluation Methods

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Use one Segment only
- Use consecutive Segments with a Hop Percentage
  Calculate the Arithmetic Mean of the Embeddings
- Use the full Utterance

Methods

Evaluation Methods

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Use one Segment only
- Use consecutive Segments with a Hop Percentage
  Calculate the Arithmetic Mean of the Embeddings
- Use the full Utterance

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Randomised Time Trajectories in Segments
- Compare different Ways to draw Segments
  - Original Time Trajectory
  - Randomised Time Trajectory of a Segment
  - Randomised Time Trajectory of the full Utterance

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- Randomised Time Trajectories in Segments
- Compare different Ways to draw Segments
  - Original Time Trajectory
  - Randomised Time Trajectory of a Segment
  - Randomised Time Trajectory of the full Utterance

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- (OT) Original Time Trajectory

- (RS) Randomised Time Trajectory of a Segment

- (RF) Randomised Time Trajectory of the full Utterance

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- (OT) Original Time Trajectory

- (RS) Randomised Time Trajectory of a Segment

- (RF) Randomised Time Trajectory of the full Utterance

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

- (OT) Original Time Trajectory

- (RS) Randomised Time Trajectory of a Segment

- (RF) Randomised Time Trajectory of the full Utterance

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

Metric		Evaluation Method		
		OT	RF	RS
Architecture	OT	OT - OT	OT - RF	OT - RS
	RF	RF - OT	RF - RF	RF - RS
	RS	RS - OT	RS - RF	RS - RS

I. A system whose results are worse or almost equal, when evaluated following OT as when following RS or RF, is not able to capture temporal context.

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

Metric		Evaluation Method		
		OT	RF	RS
Architecture	OT	OT - OT	OT - RF	OT - RS
	RF	RF - OT	RF - RF	RF - RS
	RS	RS - OT	RS - RF	RS - RS

II. A system whose results are noticeably better when evaluated following OT as when following RS, is showing first signs of being able to capture temporal context.

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

Metric		Evaluation Method		
		OT	RF	RS
Architecture	OT	OT - OT	OT - RF	OT - RS
	RF	RF - OT	RF - RF	RF - RS
	RS	RS - OT	RS - RF	RS - RS

III. A system whose results are significantly better when evaluated following OT as when following RF and RS, is showing major signs of being able to capture temporal context.

Methods

Measuring Importance of Time

School of Engineering
InIT Institut für angewandte Informationstechnologie

Metric		Evaluation Method		
		OT	RF	RS
Architecture	OT	OT - OT	OT - RF	OT - RS
	RF	RF - OT	RF - RF	RF - RS
	RS	RS - OT	RS - RF	RS - RS

IV. A system whose results are significantly better when trained and evaluated following OT as when following all other combinations of training and evaluation segment draw settings, is believed to have mastered efficiently modelling temporal context.

Results

Clean Data (TIMIT)

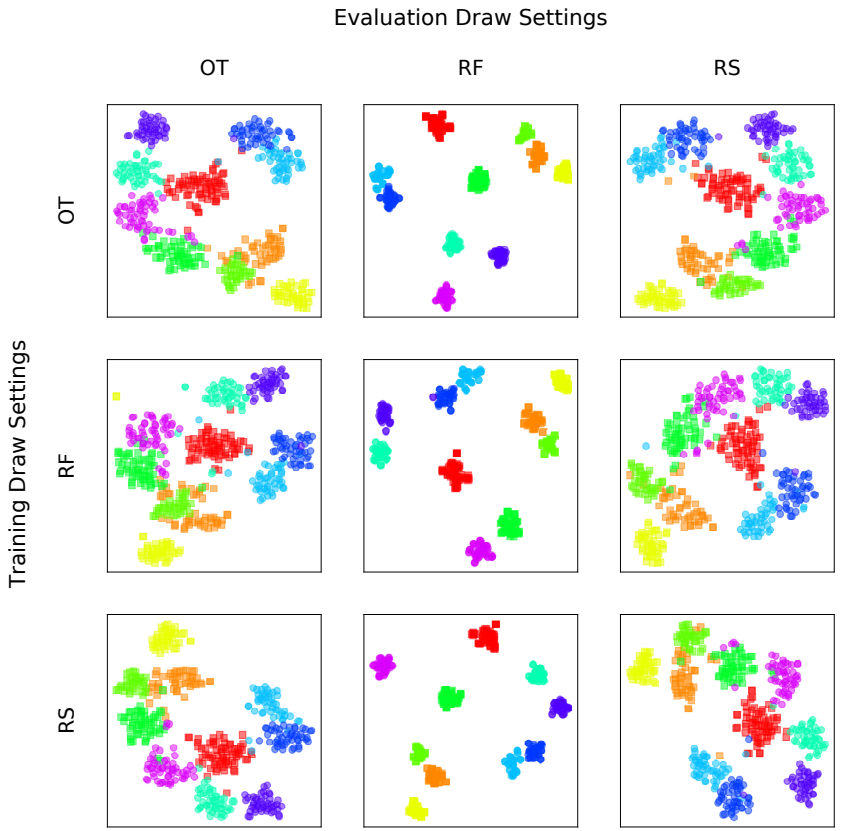
School of Engineering
InIT Institut für angewandte Informationstechnologie

EER		1 Segment			Hop 50%			Full Utterance	
		OT	RF	RS	OT	RF	RS	OT	RF/RS
MLP	OT	12.35	7.63	12.51	**6.10**	6.32	6.14	–	–
	RF	13.54	**7.32**	13.59	6.12	6.18	6.14	–	–
	RS	12.65	7.79	12.60	6.17	6.46	6.14	–	–
LUVO	OT	15.59	12.71	20.39	8.23	11.35	11.33	–	–
	RF	20.62	8.35	18.36	10.57	7.54	8.43	–	–
	RS	21.06	**7.75**	15.91	10.98	**6.87**	7.83	–	–
LSTM	OT	9.13	5.56	10.31	3.33	4.21	3.91	6.66	5.11
	RF	12.56	4.71	11.18	3.89	3.57	3.56	12.83	4.54
	RS	10.94	**4.52**	9.25	3.39	3.37	**3.18**	14.05	**4.31**
ResNet34S	OT	11.49	9.82	14.98	**5.12**	7.38	6.89	**5.30**	7.72
	RF	16.28	8.12	15.49	7.30	6.46	6.44	8.07	6.68
	RS	13.07	**7.47**	12.70	5.89	5.83	5.63	6.07	5.83
ResNet34S-TI	OT	13.90	10.64	13.90	7.43	7.83	7.43	8.34	8.34
	RF	14.63	**8.18**	14.63	5.91	**5.77**	5.91	**6.08**	**6.08**
	RS	13.99	10.58	13.99	7.08	7.51	7.08	8.19	8.19

Results

Clean Data (TIMIT)

School of Engineering
InIT Institut für angewandte Informationstechnologie



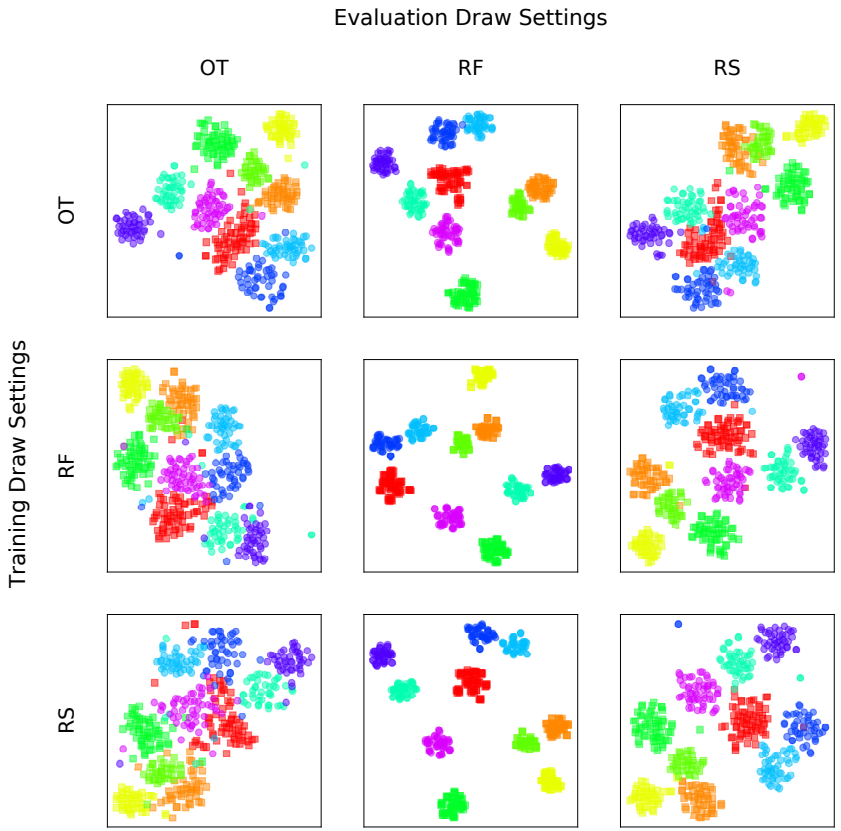
1 Segment

MLP

Results

Clean Data (TIMIT)

School of Engineering
InIT Institut für angewandte Informationstechnologie



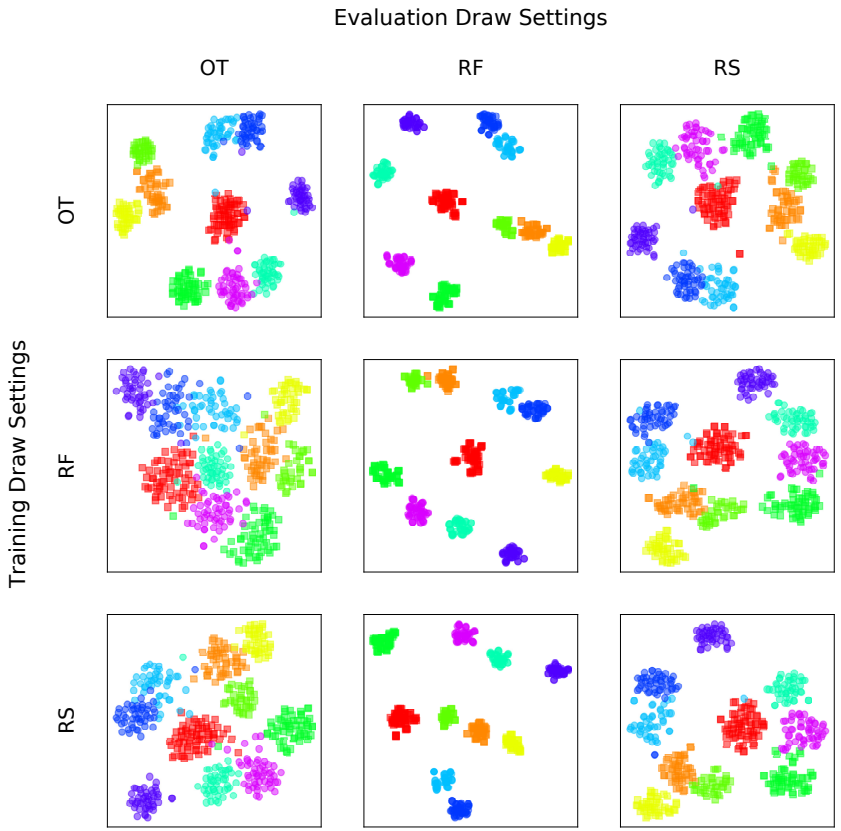
1 Segment

LUVO

Results

Clean Data (TIMIT)

School of Engineering
InIT Institut für angewandte Informationstechnologie



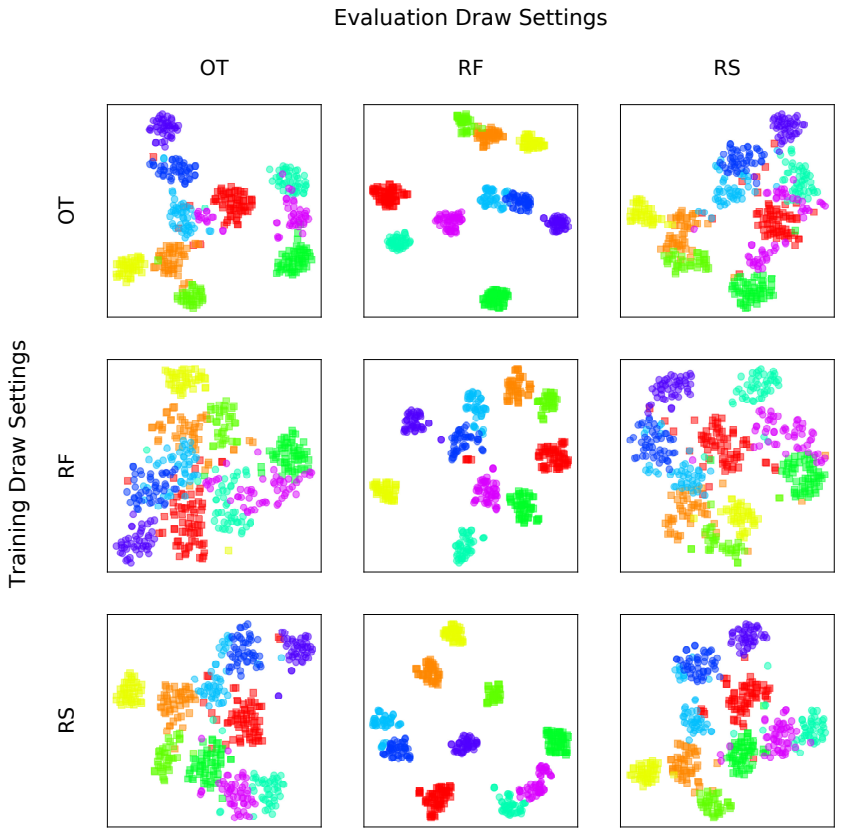
1 Segment

LSTM

Results

Clean Data (TIMIT)

School of Engineering
InIT Institut für angewandte Informationstechnologie



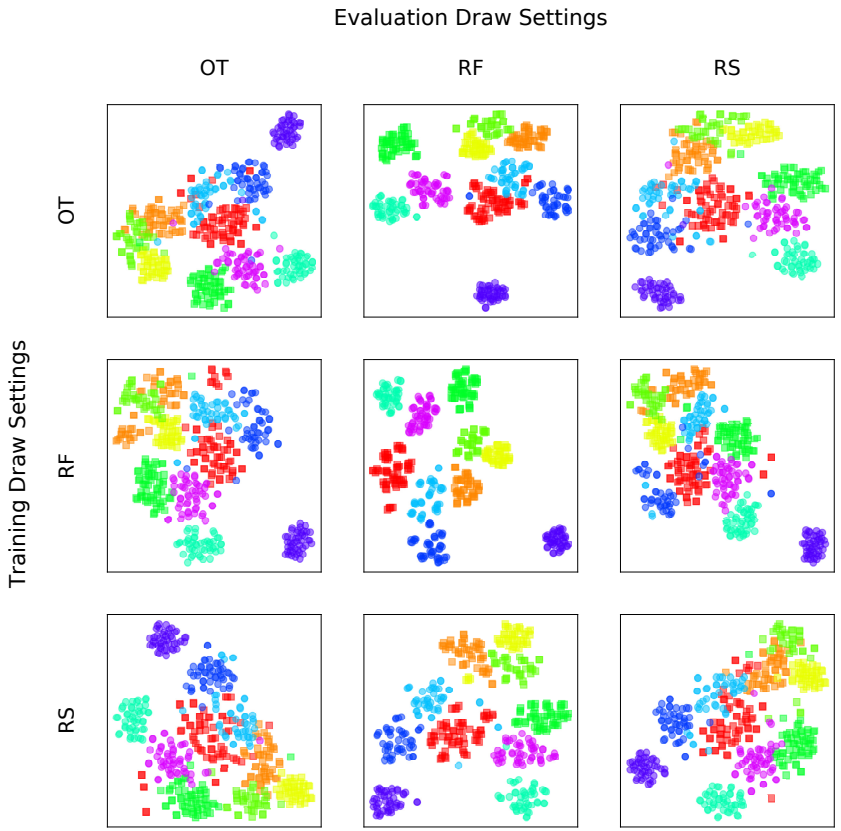
1 Segment

ResNet34S

Results

Clean Data (TIMIT)

School of Engineering
InIT Institut für angewandte Informationstechnologie



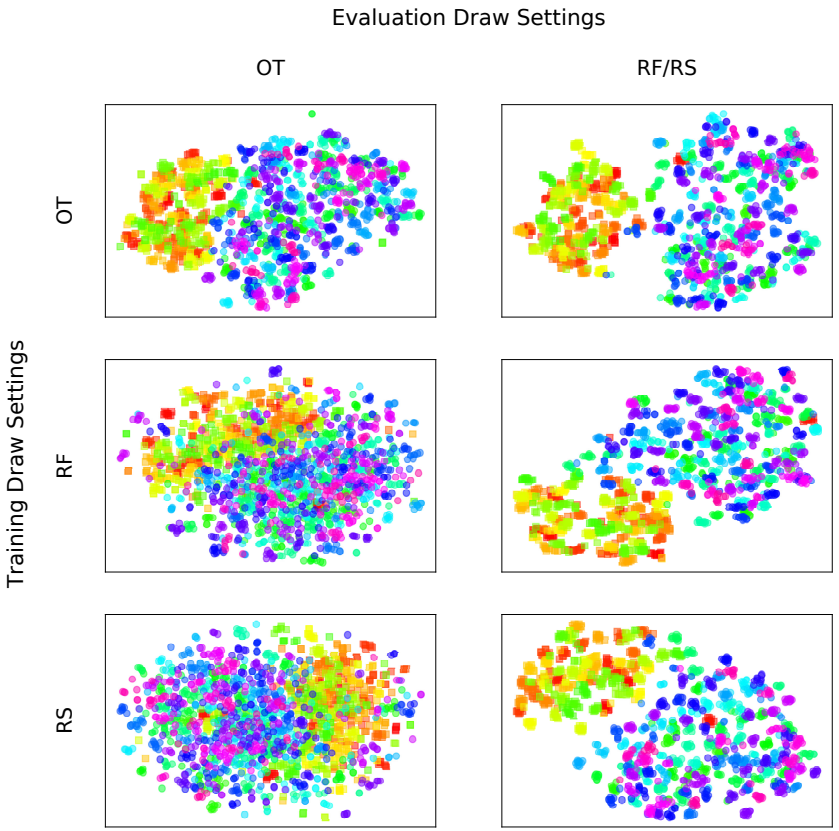
1 Segment

ResNet34S - TI

Results

Clean Data (TIMIT)

School of Engineering
InIT Institut für angewandte Informationstechnologie



Full Utterance

LSTM

Results

Real World Data (VoxCeleb)

School of Engineering
InIT Institut für angewandte Informationstechnologie

EER		1 Segment			Hop 50%			Full Utterance	
		OT	RF	RS	OT	RF	RS	OT	RF/RS
LSTM	OT	16.81	22.15	26.56	13.31	21.70	21.17	**11.47**	21.50
	RF	21.76	**13.65**	20.84	14.39	**12.53**	13.23	17.38	13.77
	RS	24.26	19.33	22.94	20.72	18.85	19.22	21.00	18.63
ResNet34S	OT	17.59	28.91	33.22	**8.94**	24.84	22.52	**8.47**	25.85
	RF	28.69	18.11	27.28	18.00	15.97	15.86	17.44	16.43
	RS	22.89	**17.42**	23.48	16.59	15.49	15.48	15.72	15.84

Conclusions

School of Engineering
InIT Institut für angewandte Informationstechnologie

- What have we learned?
- Future Work
- Personal take-away

Conclusions

School of Engineering
InIT Institut für angewandte Informationstechnologie

- What have we learned?
- Future Work
- Personal take-away

Conclusions

School of Engineering
InIT Institut für angewandte Informationstechnologie

- What have we learned?
- Future Work
- Personal take-away

Questions?

School of Engineering
InIT Institut für angewandte Informationstechnologie