Using Past Speaker Behavior to Better Predict Turn Transitions

Thesis Presentation

Tomer Meshorer

Center for Spoken Language Understanding
Oregon Health & Science University
Portland, Oregon,USA

08 June 2017

Outline

1 Motivation

2 Theoretical Model

3 Data Preparation

4 Data Exploration

5 Machine Learning Models

6 Summary

‘

Section 1

Motivation

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment
   - too little - machine barge in during intra turn pause.

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment
   - too little - machine barge in during intra turn pause.
   - too much - user waiting for the machine.

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment
   - too little - machine barge in during intra turn pause.
   - too much - user waiting for the machine.
4. Turn transition prediction based on local features improve turn taking but still do not match human performance.

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment
   - too little - machine barge in during intra turn pause.
   - too much - user waiting for the machine.
4. Turn transition prediction based on local features improve turn taking but still do not match human performance.
   - Syntactic (Sacks et al 1978,De Ruiter et al. 2006)

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment
   - too little - machine barge in during intra turn pause.
   - too much - user waiting for the machine.
4. Turn transition prediction based on local features improve turn taking but still do not match human performance.
   - Syntactic (Sacks et al 1978,De Ruiter et al. 2006)
   - Prosodic (Ford 1996,Stolcke 2002,Ferrer 2003)

Current Issues

1. For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
2. In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
3. Timeouts leads to poor user interaction(Arsikere et al, 2015)
   - Not effective in noisy environment
   - too little - machine barge in during intra turn pause.
   - too much - user waiting for the machine.
4. Turn transition prediction based on local features improve turn taking but still do not match human performance.
   - Syntactic (Sacks et al 1978,De Ruiter et al. 2006)
   - Prosodic (Ford 1996,Stolcke 2002,Ferrer 2003)
   - Pragmatic (Ford 2001)

Thesis Statement

Conversant's past behavior can help predict turn transitions

Past behavior represented by Summary features

Acknowledgement

1. This work was partially funded by the National Science Foundation under grant IIS-1321146.

Acknowledgement

1. This work was partially funded by the National Science Foundation under grant IIS-1321146.
2. This thesis is based on a paper that was submitted and presented at Interspeech 2016.

Acknowledgement

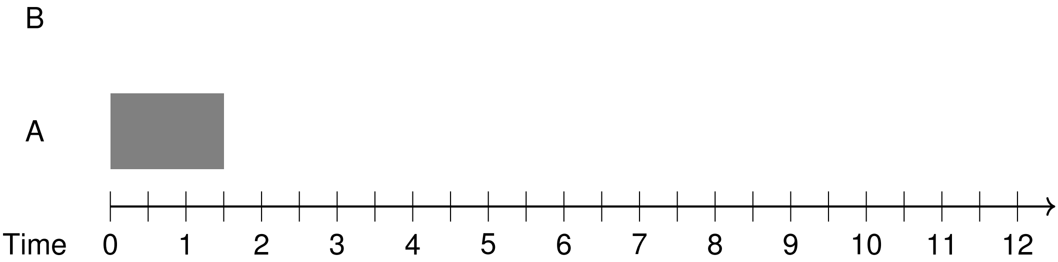
1. This work was partially funded by the National Science Foundation under grant IIS-1321146.
2. This thesis is based on a paper that was submitted and presented at Interspeech 2016.
3. Thesis advisor and collaborator : Prof. Peter Heeman.

Section 2

Theoretical Model

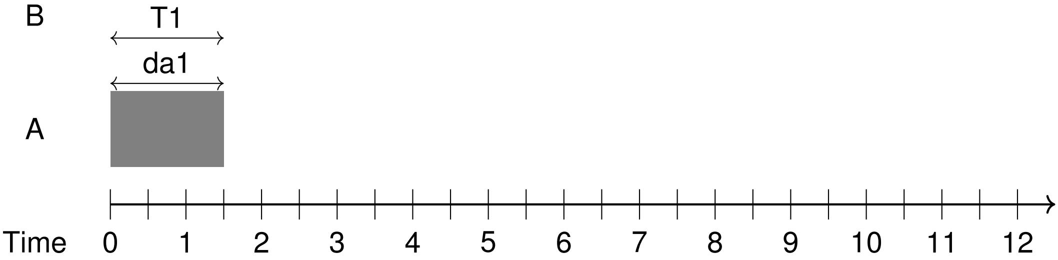
Relative Turn Length

Timing Diagram



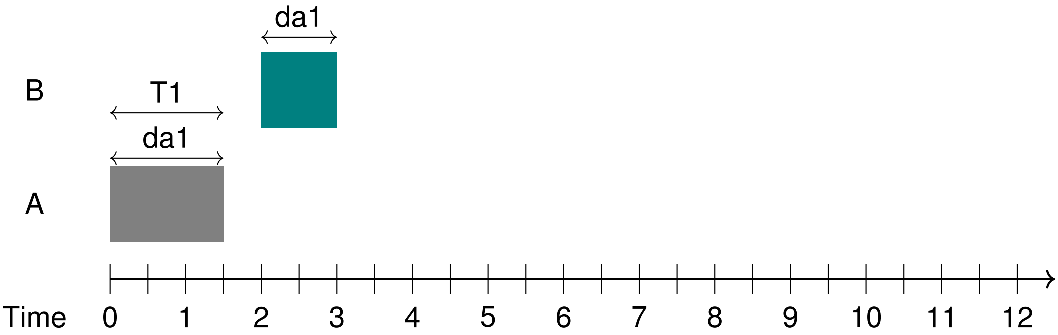
Relative Turn Length

Timing Diagram



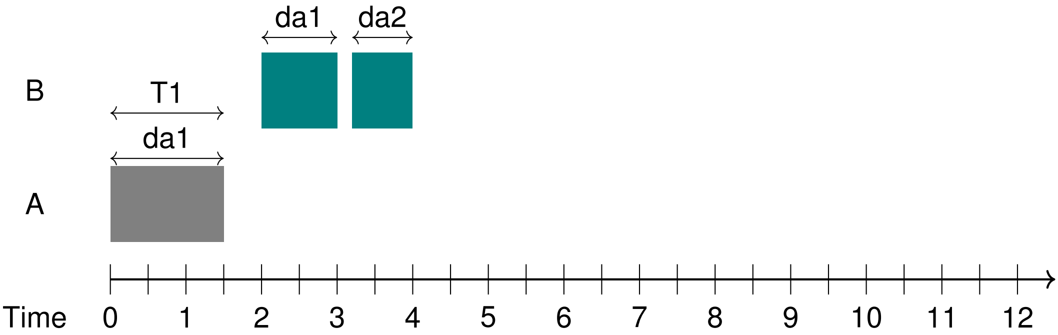
Relative Turn Length

Timing Diagram



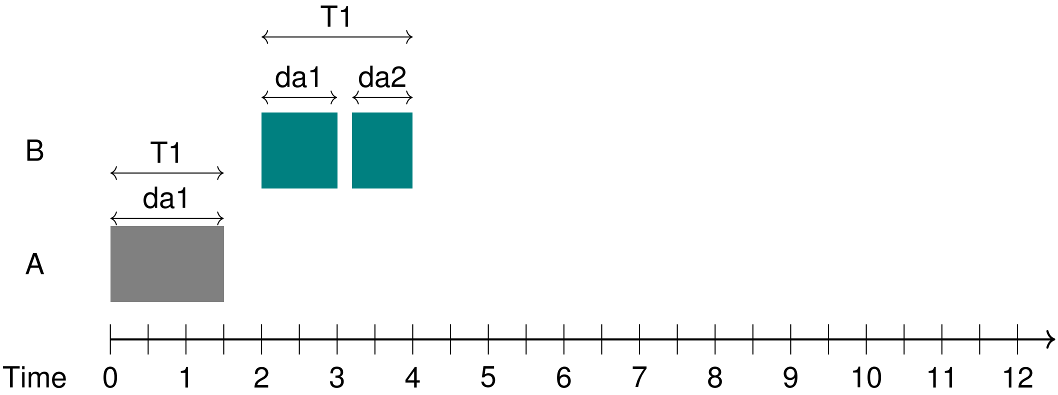
Relative Turn Length

Timing Diagram



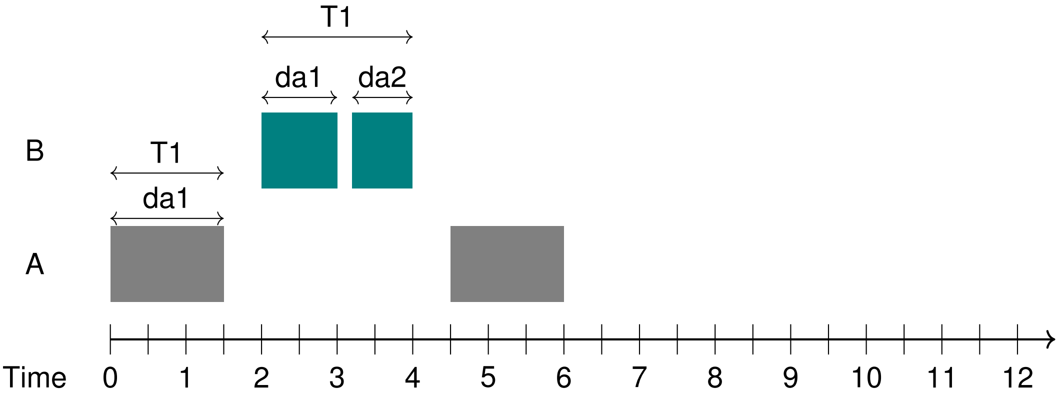
Relative Turn Length

Timing Diagram



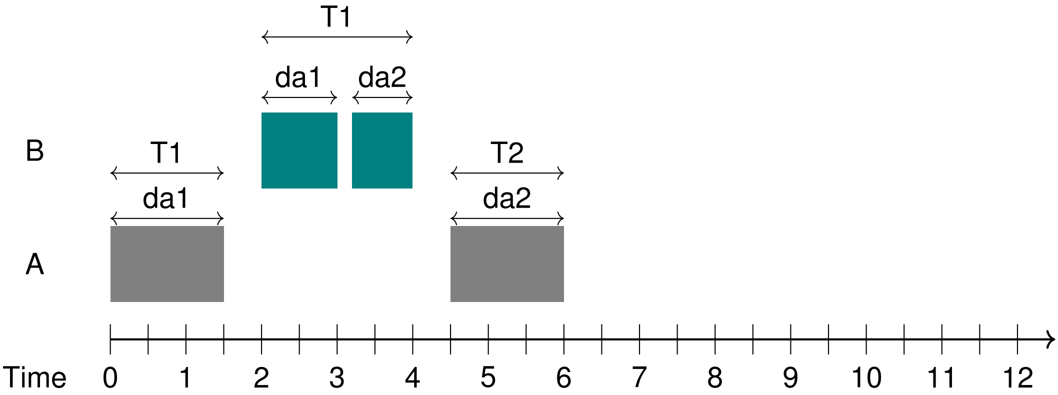
Relative Turn Length

Timing Diagram



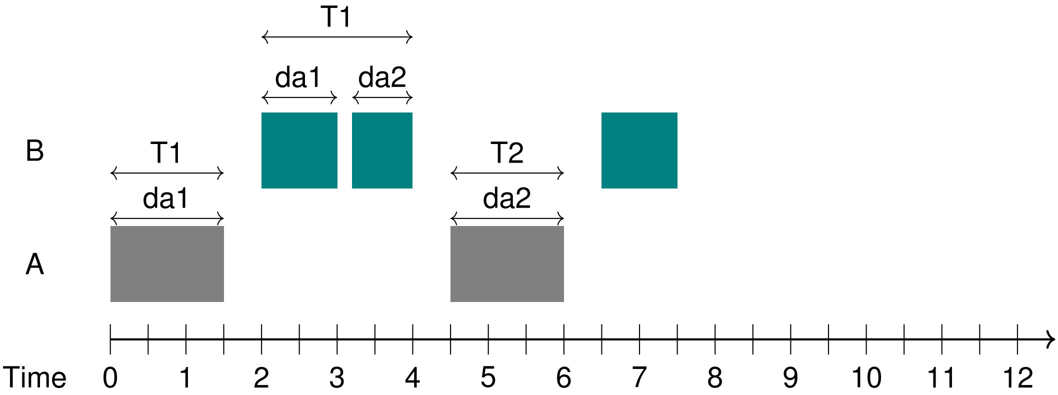
Relative Turn Length

Timing Diagram



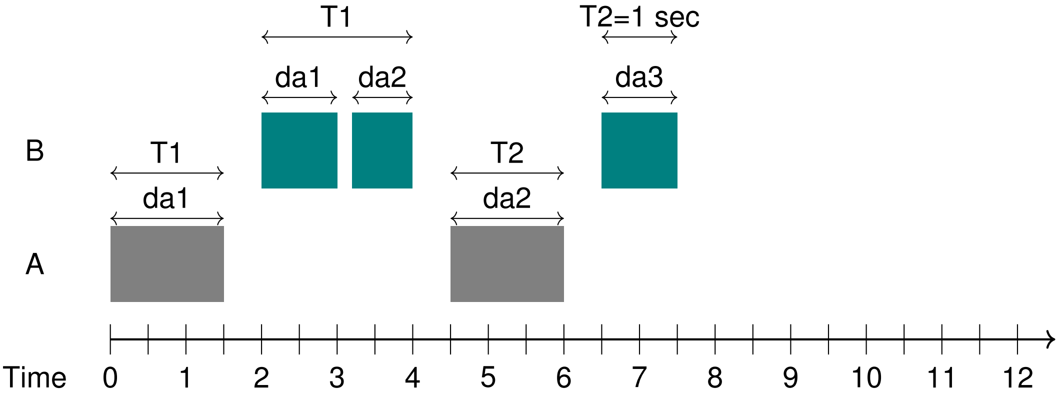
Relative Turn Length

Timing Diagram



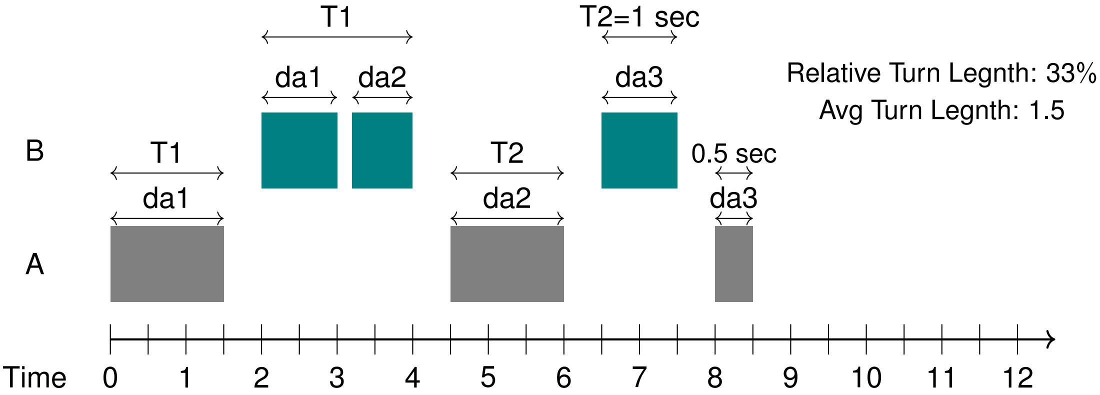
Relative Turn Length

Timing Diagram



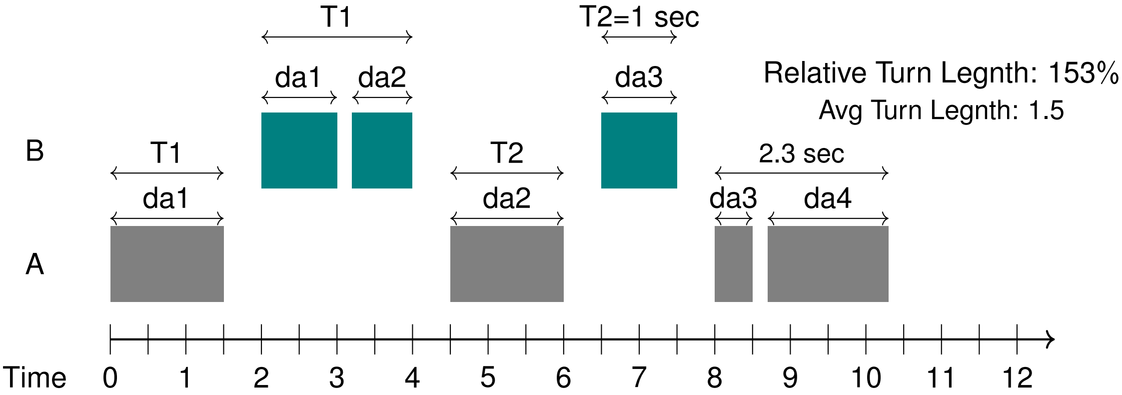
Relative Turn Length

Timing Diagram



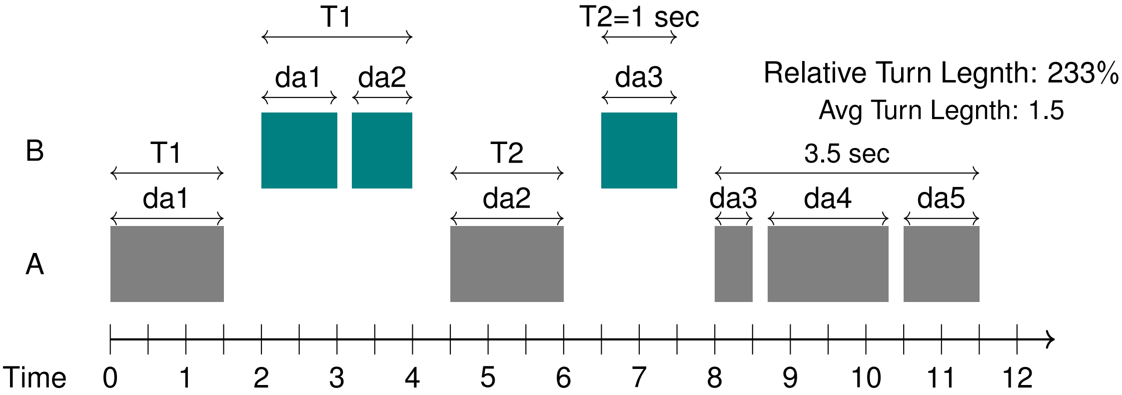
Relative Turn Length

Timing Diagram



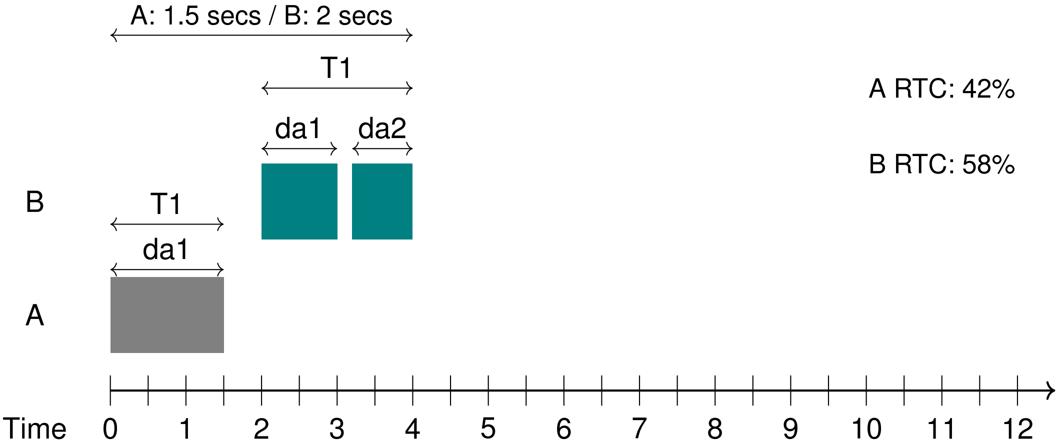
Relative Turn Length

Timing Diagram



Relative Floor Control

Timing Diagram

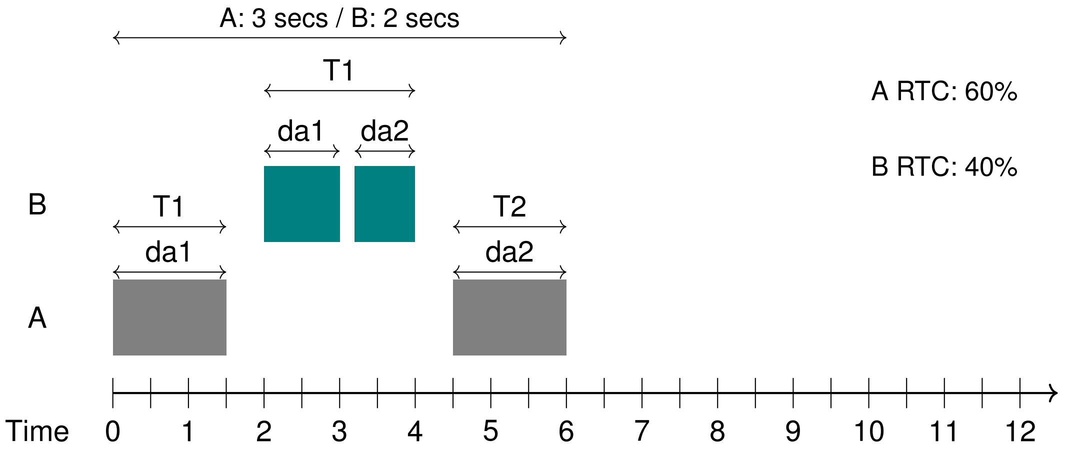


A RTC: 42%

B RTC: 58%

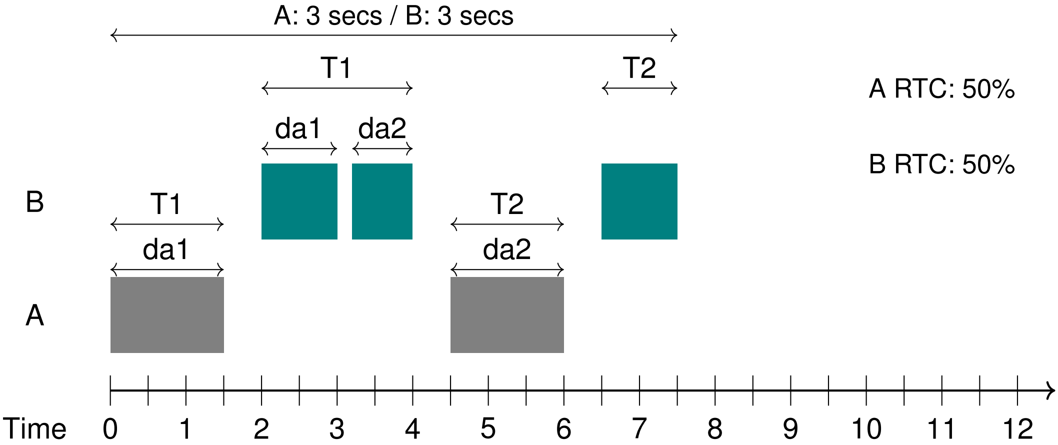
Relative Floor Control

Timing Diagram



Relative Floor Control

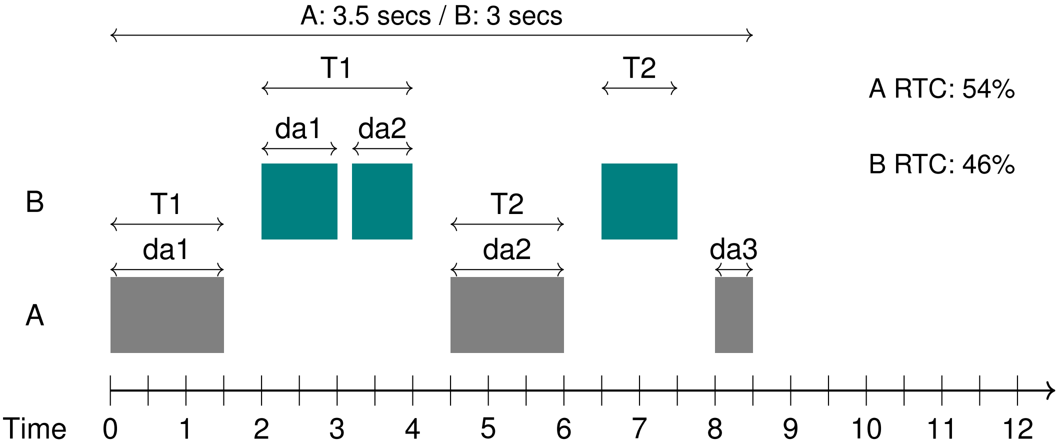
Timing Diagram



A RTC: 50%

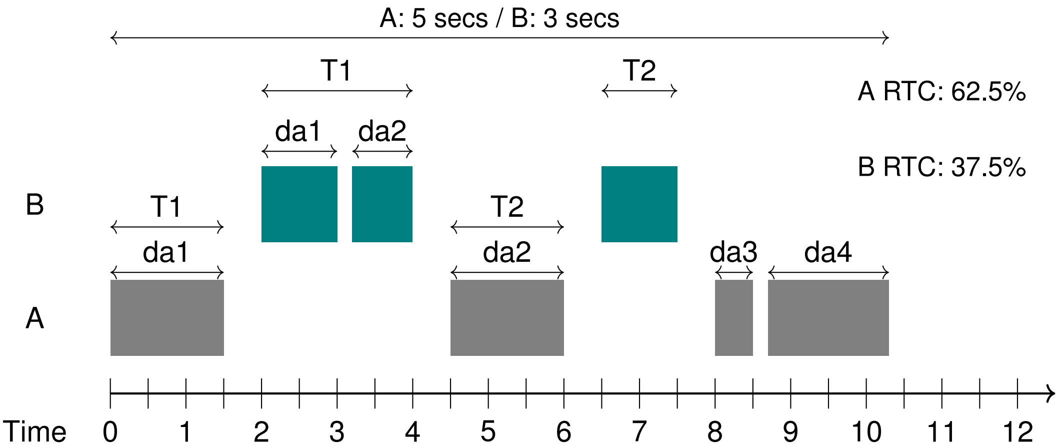
B RTC: 50%

Relative Floor Control
Timing Diagram



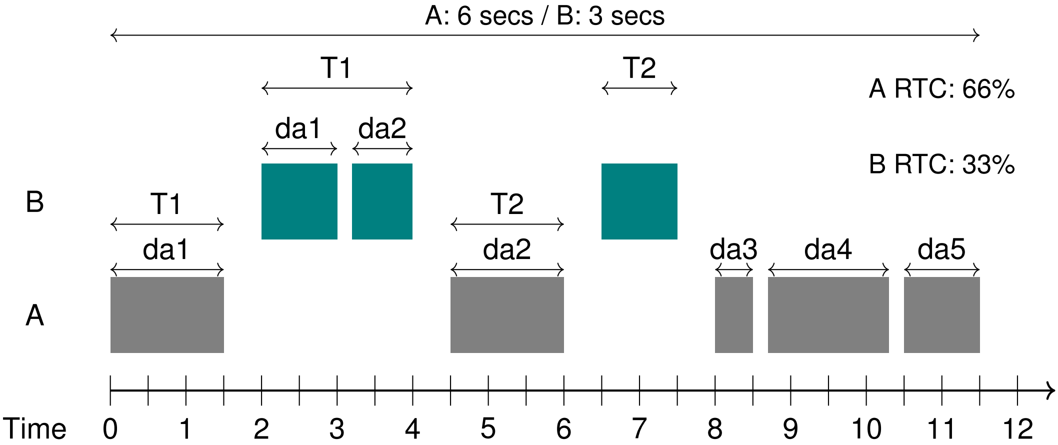
Relative Floor Control

Timing Diagram



Relative Floor Control

Timing Diagram



‘

Section 3

Data Preparation

Corpus

1. Switchboard corpus (NXT).

Corpus

1. Switchboard corpus (NXT).
2. Audio recording of casual conversations between randomly chosen speakers.

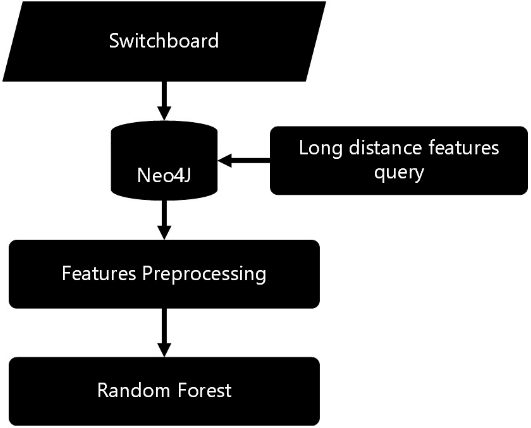
Corpus

1. Switchboard corpus (NXT).
2. Audio recording of casual conversations between randomly chosen speakers.
3. 2483 conversation, involving 520 speakers

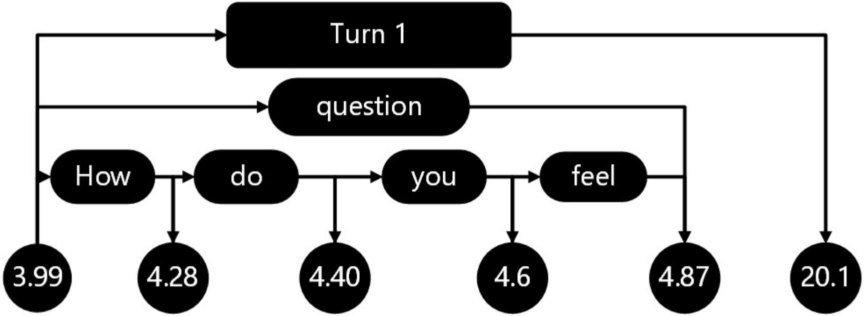
Corpus

1. Switchboard corpus (NXT).
2. Audio recording of casual conversations between randomly chosen speakers.
3. 2483 conversation, involving 520 speakers
4. In our research the corpus contain 642 conversations.

Preprocessing pipeline



Conversation representation



Preprocessing

- Removed 11 dialogue acts that were coded as other in switchboard.
- Skip the first 120 seconds of the conversation.
  - Gives time for conversant to form the conversional image.
  - Reduces the dialogue acts from 50633 to 37508.
- Reduce data sparsity by collapsing 65 dialog acts into 9.

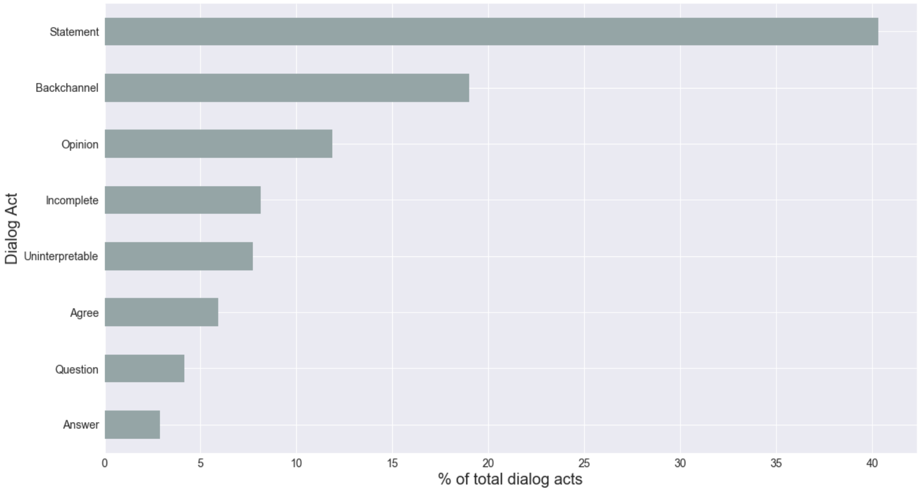
Switchboard dialog acts	Dialog act classes
sd,h,bf	statement
sv,ad,sv@	statement - opinion
aa,aar̂	agree accept
%.%-,%@	abandon
b,bh	backchannel
qy,qo,qh	question
no,ny,ng,arp	answer
+	+
o@,+@	NA

Table: Mapping from dialog act to dialog act class

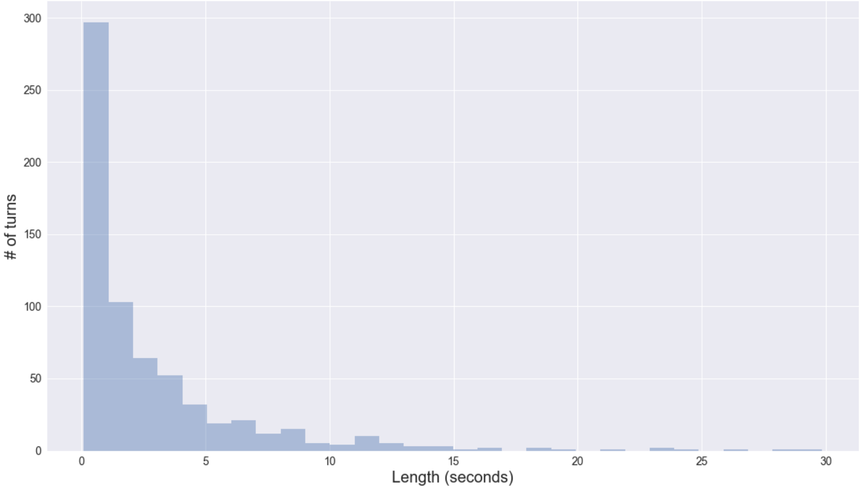
Section 4

Data Exploration

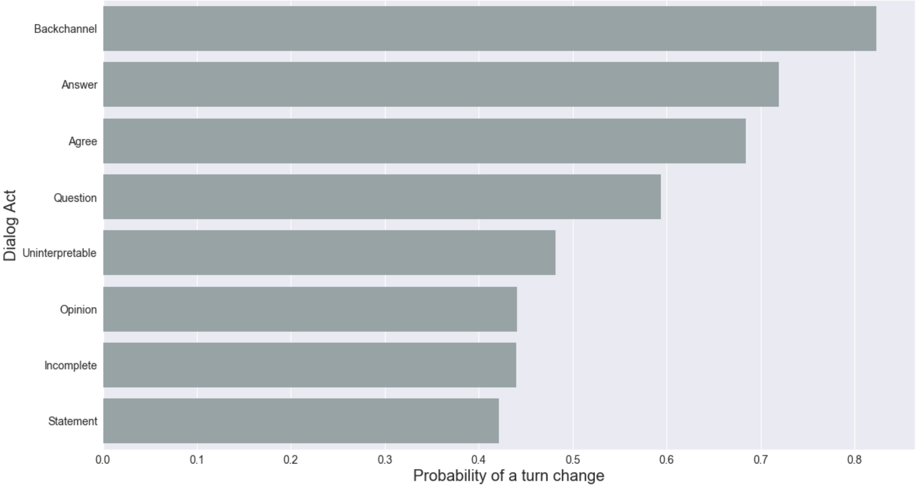
Dialog act relative count



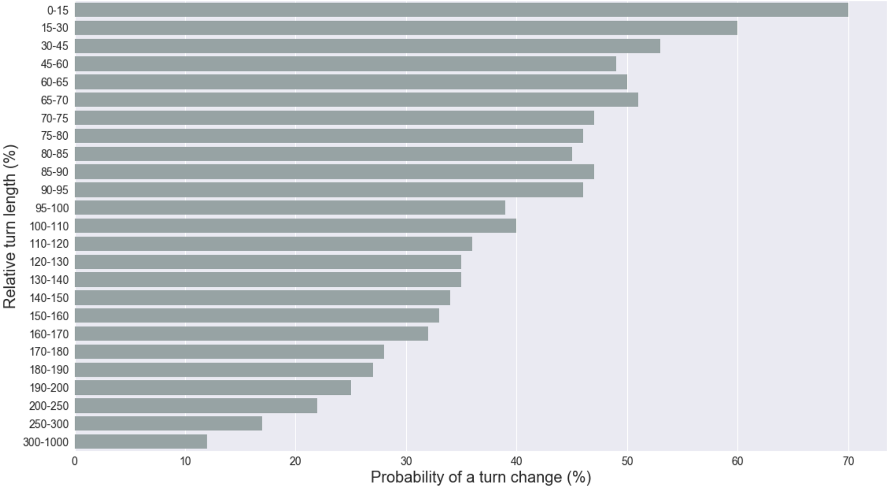
Turn Length Distribution



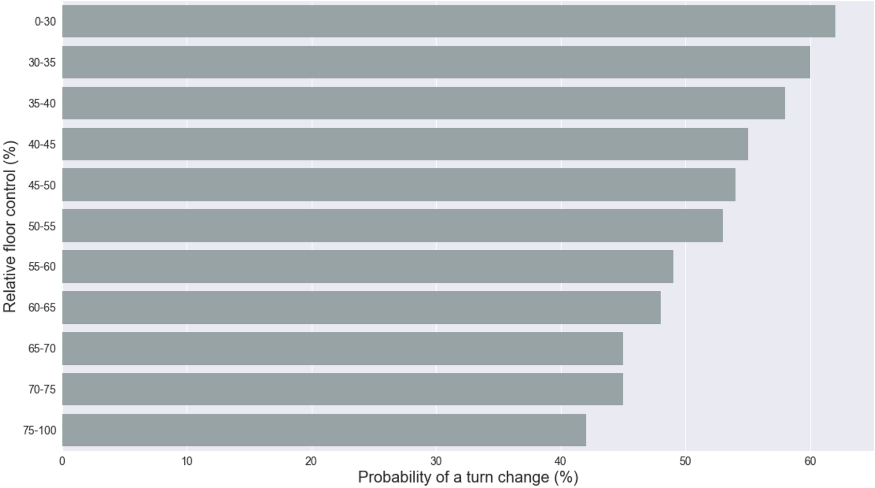
Dialog act probability of turn change



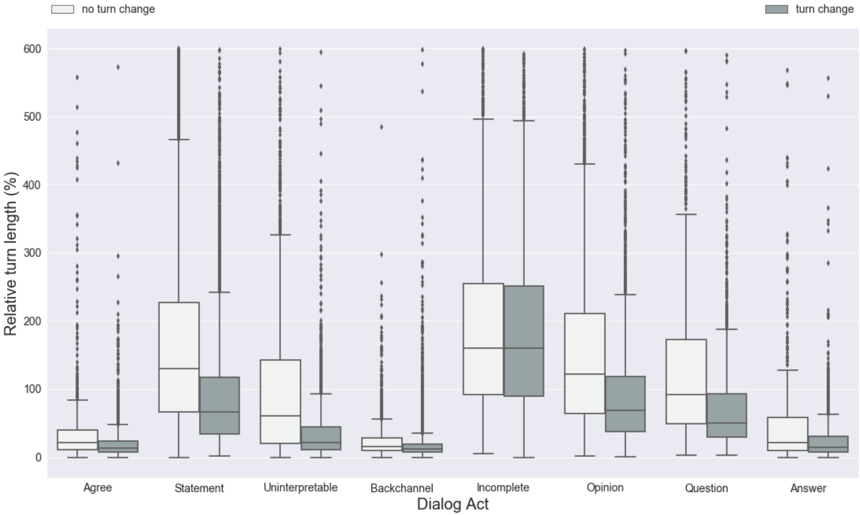
Relative Turn Length effect on probability of turn change



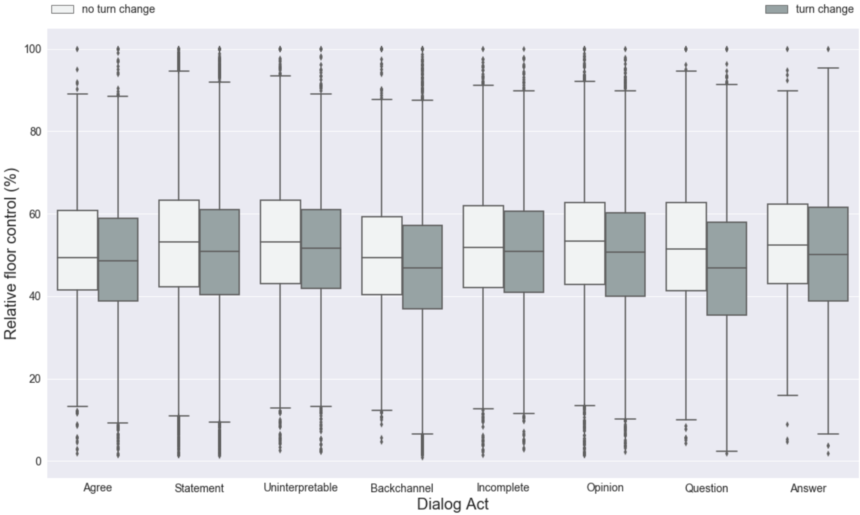
Relative Turn Control effect on probability of a turn change



Relative turn length for dialog act type



Relative floor control by dialog act



‘

Section 5

Machine Learning Models

Classifiers

- Used random forests (N=200) / Gradient Boosting to train and test the following models
  - baseline 1: current dialog act label.
  - baseline 2: current and previous dialog acts.
  - summary model: just the summary features.
  - full model: summary features and the current and previous dialog acts.
- Evaluation was done using 10 fold cross validation.
- Run grid search to find the optimal hyper parameters.

Result for Random Forest Classifier

	Accuracy	F1	Precision	Recall	AUC
baseline 1	62.79%	57.81%	74.98%	47.04%	65.99%
baseline 2	74.89%	74.87%	81.84%	69.00%	81.11%
summary	65.54%	69.32%	67.22%	71.36%	69.46%
full	75.75%	77.59%	77.50%	77.83%	83.78%

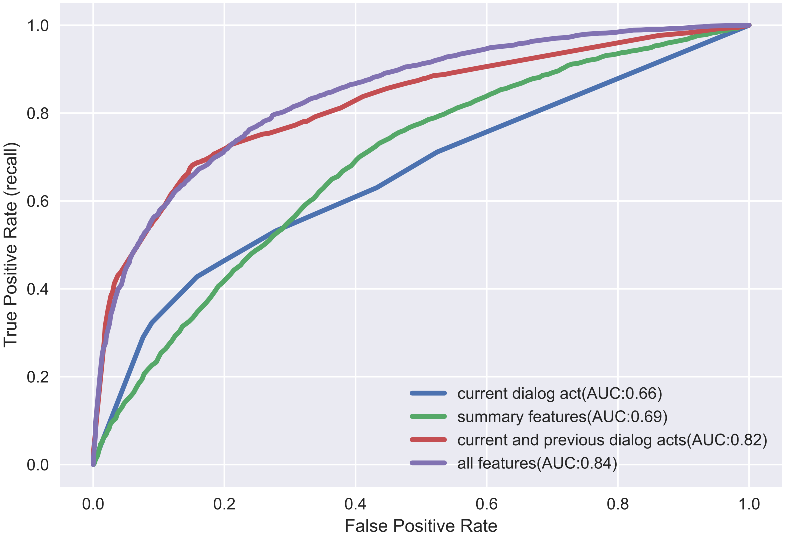
Table: Precision, recall and F1 results using Random Forests

Result for Gradient Boosting

	Accuracy	F1	Precision	Recall	AUC
baseline 1	62.79%	57.81%	74.98%	47.04%	65.99%
baseline 2	74.88%	74.82%	81.92%	68.86%	81.10%
summary	67.91%	71.30%	69.20%	73.55%	72.64%
all	76.57%	78.74%	77.44%	80.11%	84.84%

Table: Precision, recall and F1 results using Gradient boost classifier

ROC curves and AUC of different models



Sensitivity to Measurement Start Time

	0s	15s	30s	45s	60s	**120s**	180s
baseline 1	65.99%	66.10%	66.12%	66.09%	66.02%	65.98%	66.05%
baseline 2	81.11%	81.21%	81.24%	81.20%	81.15%	80.92%	80.68%
summary	69.46%	69.51%	69.43%	69.49%	69.57%	69.10%	69.21%
full	83.78%	83.87%	83.85%	83.80%	83.61%	83.19%	82.80%

Table: AUC Score in relation to the start of the dialog

Section 6

Summary

Conclusion

1. Summary features do provide improvement over local features.

Conclusion

1. Summary features do provide improvement over local features.
2. However, the affect for our data is the opposite of our initial assumption

Conclusion

1. Summary features do provide improvement over local features.
2. However, the affect for our data is the opposite of our initial assumption
   - Short turn (Low RTL) leads to turn change

Conclusion

1. Summary features do provide improvement over local features.
2. However, the affect for our data is the opposite of our initial assumption
   - Short turn (Low RTL) leads to turn change
   - In long turn the speaker will actually hold the floor.

Future work

1. Combine the summary features with other local features (semantic/prosadic)

Future work

1. Combine the summary features with other local features (semantic/prosadic)
2. Test the hypothesis on other corpus (for example task based corpus

Future work

1. Combine the summary features with other local features (semantic/prosadic)
2. Test the hypothesis on other corpus (for example task based corpus
3. Instead of measuring the affect from the start of the conversation, use moving averages with different window length.

Future work

1. Combine the summary features with other local features (semantic/prosadic)
2. Test the hypothesis on other corpus (for example task based corpus
3. Instead of measuring the affect from the start of the conversation, use moving averages with different window length.
4. Perform the experiments where back channels are not considered as turn change.

Future work

1. Combine the summary features with other local features (semantic/prosadic)
2. Test the hypothesis on other corpus (for example task based corpus
3. Instead of measuring the affect from the start of the conversation, use moving averages with different window length.
4. Perform the experiments where back channels are not considered as turn change.
5. In general, any local features can be turned into a summary feature by taking the avarage over past turn. Hence this area of research can be expanded to other local features.