Analysis of the Occurrence of Laughter in Meetings

Kornel Laskowski[1,2] & Susanne Burger[2]

[1]interACT, Universität Karlsruhe
[2]interACT, Carnegie Mellon University

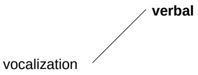
August 29, 2007

Introduction

- primary motivation: meeting understanding

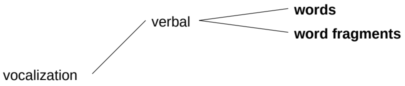
Introduction

- primary motivation: meeting understanding



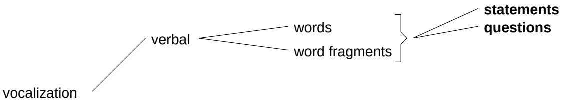
Introduction

- primary motivation: meeting understanding



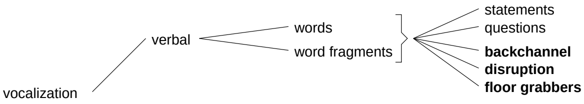
Introduction

- primary motivation: meeting understanding



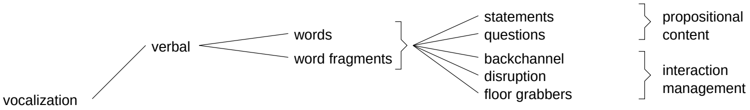
Introduction

- primary motivation: meeting understanding



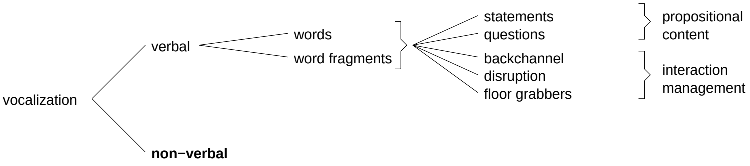
Introduction

- primary motivation: meeting understanding



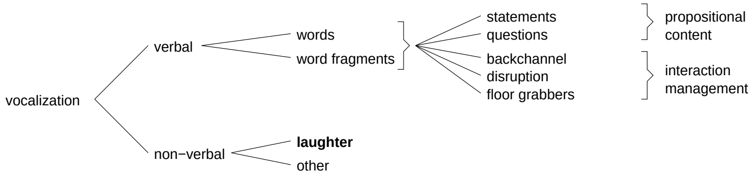
Introduction

- primary motivation: meeting understanding



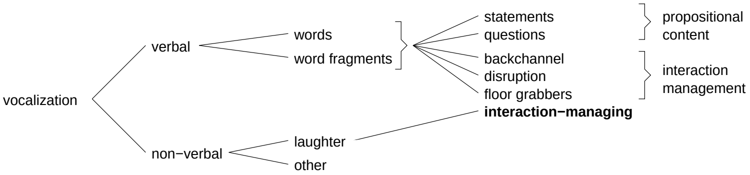
Introduction

- primary motivation: meeting understanding



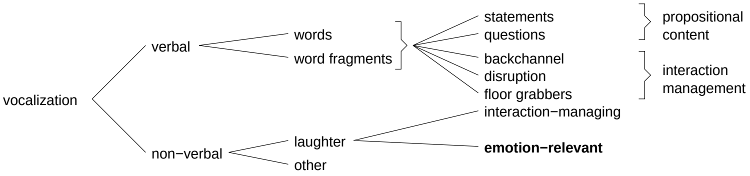
Introduction

- primary motivation: meeting understanding



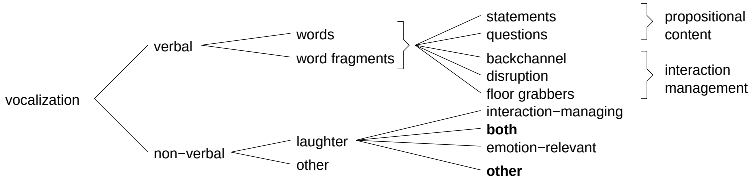
Introduction

- primary motivation: meeting understanding



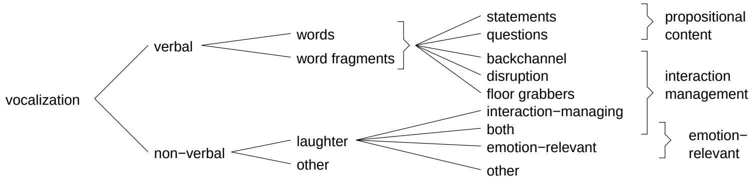
Introduction

- primary motivation: meeting understanding



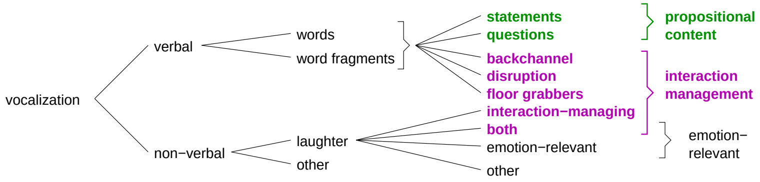
Introduction

- primary motivation: meeting understanding



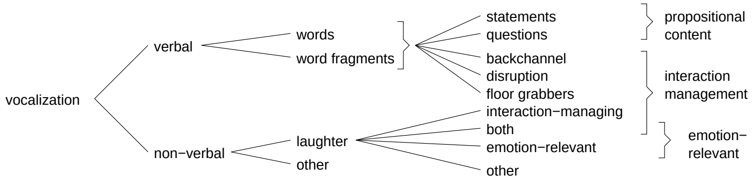
Introduction

- primary motivation: meeting understanding



Introduction

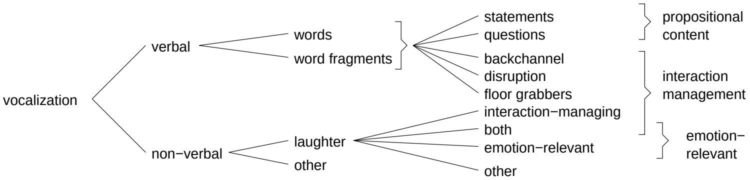
- primary motivation: meeting understanding



- laughter detection is particularly important for understanding both interaction and emotion **if laughter occurs frequently**

Introduction

- primary motivation: meeting understanding



- laughter detection is particularly important for understanding both interaction and emotion **if laughter occurs frequently**
- to date, for meetings, it is not known
  1. how much laughter there actually is
  2. when it tends to occur

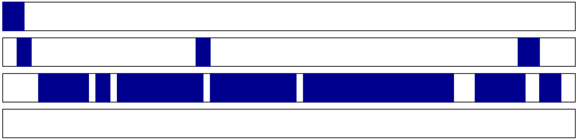
Text-Independent Modeling of Multi-Participant Meetings

To find interaction, model participants **jointly**.

Text-Independent Modeling of Multi-Participant Meetings

To find interaction, model participants **jointly**.

- essentially monologue



Text-Independent Modeling of Multi-Participant Meetings

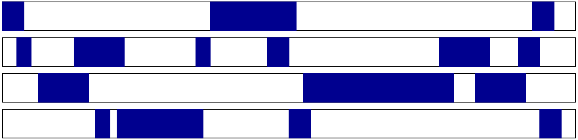
To find interaction, model participants **jointly**.

- "multi-logue"

Text-Independent Modeling of Multi-Participant Meetings

To find interaction, model participants **jointly**.

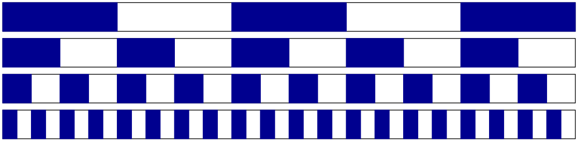
- “multi-logue” with more participant involvement



Text-Independent Modeling of Multi-Participant Meetings

To find interaction, model participants **jointly**.

- a mathematical artifact (the Haar wavelet basis)



Text-Independent Modeling of Multi-Participant Meetings

To find interaction, model participants **jointly**.

- “multi-logue”

Text-Independent Modeling of Multi-Participant Meetings

To find interaction, model participants **jointly**.

- “multi-logue” with laughter
  - participants tend to wait to speak
  - participants **do not** wait to laugh

Three Questions of Interest

1. What is the quantity of laughter, relative to the quantity of speech?

Three Questions of Interest

1. What is the quantity of laughter, relative to the quantity of speech?
2. How does the durational distribution of episodes of laughter differ from that of episodes of speech?

Three Questions of Interest

1. What is the quantity of laughter, relative to the quantity of speech?
2. How does the durational distribution of episodes of laughter differ from that of episodes of speech?
3. How do meeting participants appear to affect each other in their use of laughter, relative to their use of speech?

Laugh Bouts vs Talk Spurts

- we will contrast the occurrence of laughter $\mathcal{L}$ with that of speech $\mathcal{S}$

Laugh Bouts vs Talk Spurts

- we will contrast the occurrence of laughter $\mathcal{L}$ with that of speech $\mathcal{S}$

talk spurts: contiguous per-participant intervals of speech (Shriberg et al, 2001), containing pauses no longer than 300 ms (as in NIST RT-06s SAD)

Laugh Bouts vs Talk Spurts

- we will contrast the occurrence of laughter $\mathcal{L}$ with that of speech $\mathcal{S}$

talk spurts contiguous per-participant intervals of speech (Shriberg et al, 2001), containing pauses no longer than 300 ms (as in NIST RT-06s SAD)

laugh bouts contiguous per-participant intervals of laughter (Bachorowski et al, 2001), including recovery inhalation

Laugh Bouts vs Talk Spurts

- we will contrast the occurrence of laughter $\mathcal{L}$ with that of speech $\mathcal{S}$

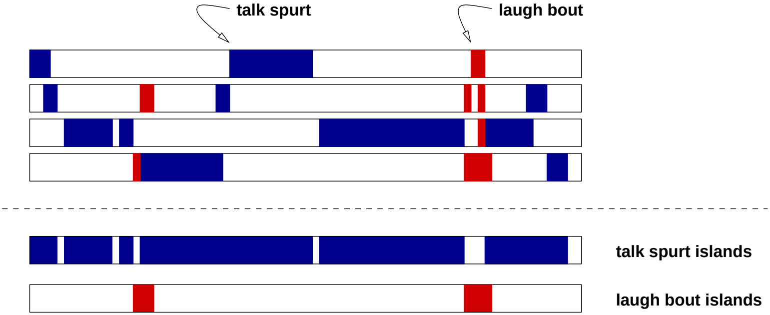
talk spurts contiguous per-participant intervals of speech (Shriberg et al, 2001), containing pauses no longer than 300 ms (as in NIST RT-06s SAD)

laugh bouts contiguous per-participant intervals of laughter (Bachorowski et al, 2001), including recovery inhalation

$\mathcal{S}/\mathcal{L}$ islands contiguous per-group intervals in which at least one participant talks/laughs

Laugh Bouts vs Talk Spurts

- we will contrast the occurrence of laughter $\mathcal{L}$ with that of speech $\mathcal{S}$



The ICSI Meeting Corpus

- naturally occurring project-oriented conversations with varying number of participants

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations with varying number of participants
- the largest such corpus available

type	# of meetings	# of participants		
		mod	min	max
Bed	15	6	4	7
Bmr	29	7	3	9
Bro	23	6	4	8
other	8	6	5	8

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations with varying number of participants
- the largest such corpus available

type	# of meetings	# of participants		
		mod	min	max
Bed	15	6	4	7
Bmr	29	7	3	9
Bro	23	6	4	8
other	8	6	5	8

- rarely, meetings contain additional, uninstrumented participants (we ignore them)

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations with varying number of participants
- the largest such corpus available

type	# of meetings	# of participants		
		mod	min	max
Bed	15	6	4	7
Bmr	29	7	3	9
Bro	23	6	4	8
other	8	6	5	8

- rarely, meetings contain additional, uninstrumented participants (we ignore them)
- we use all 75 meetings: 66.3 hours of conversation

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:
  1. identifying laughter in the orthographic transcription

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:
  1. identifying laughter in the orthographic transcription
  2. specifying endpoints for identified laughter

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:
  1. identifying laughter in the orthographic transcription
  2. specifying endpoints for identified laughter

1. orthographic, time-segmented transcription of speaker contributions (`.stm`)

```
Bmr011 me013 chan1 3029.466 3029.911 Yeah.
Bmr011 mn005 chan3 3030.230 3031.140 Film-maker.
Bmr011 fe016 chan0 3030.783 3032.125 <Emphasis> colorful. </Emphasi...
Bmr011 me011 chanB 3035.301 3036.964 Of beeps, yeah.
Bmr011 fe008 chan8 3035.714 3037.314 <Pause/> of m- one hour of - <...
Bmr011 mn014 chan2 3036.030 3036.640 Yeah.
Bmr011 me013 chan1 3036.280 3037.600 <VocalSound Description="laugh"/>
Bmr011 mn014 chan2 3036.640 3037.115 Yeah.
Bmr011 mn005 chan3 3036.930 3037.335 Is -
Bmr011 me011 chanB 3036.964 3038.573 <VocalSound Description="laugh"/>
```

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:
  1. identifying laughter in the orthographic transcription
  2. specifying endpoints for identified laughter

1. orthographic, time-segmented transcription of speaker contributions (`.stm`)

```
...9.911 Yeah.
...1.140 Film-maker.
...2.125 <Emphasis> colorful. </Emphasis> <Comment Description="while laughing"/>
...6.964 Of beeps, yeah.
...7.314 <Pause/> of m- one hour of - <Comment Description="while laughing"/>
...6.640 Yeah.
...7.600 <VocalSound Description="laugh"/>
...7.115 Yeah.
...7.335 Is -
...8.573 <VocalSound Description="laugh"/>
```

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:
  1. identifying laughter in the orthographic transcription
  2. specifying endpoints for identified laughter

1. orthographic, time-segmented transcription of speaker contributions (`.stm`)

```
...9.911 Yeah.
...1.140 Film-maker.
...2.125 <Emphasis> colorful. </Emphasis> <Comment Description="while laughing"/>
...6.964 Of beeps, yeah.
...7.314 <Pause/> of m- one hour of - <Comment Description="while laughing"/>
...6.640 Yeah.
...7.600 <VocalSound Description="laugh"/>
...7.115 Yeah.
...7.335 Is -
...8.573 <VocalSound Description="laugh"/>
```

Identifying Laughter in the ICSI Corpus

- laughter is already annotated with rich XML-style mark-up
- therefore, for our purposes, data preprocessing consists of:
  1. identifying laughter in the orthographic transcription
  2. specifying endpoints for identified laughter

1. orthographic, time-segmented transcription of speaker contributions (`.stm`)

```
...9.911 Yeah.
...1.140 Film-maker.
...2.125 <Emphasis> colorful. </Emphasis> <Comment Description="while laughing"/>
...6.964 Of beeps, yeah.
...7.314 <Pause/> of m- one hour of - <Comment Description="while laughing"/>
...6.640 Yeah.
...7.600 <VocalSound Description="laugh"/>
...7.115 Yeah.
...7.335 Is -
...8.573 <VocalSound Description="laugh"/>
```

Sample VocalSound Instances

Freq Rank	Token Count	VocalSound Description	Used
1	11515	laugh	$\surd$
2	7091	breath	
3	4589	inbreath	
4	2223	mouth	
5	970	breath-laugh	$\surd$
11	97	laugh-breath	$\surd$
46	6	cough-laugh	$\surd$
63	3	laugh, "hmmph"	$\surd$
69	3	breath while smiling	
75	2	very long laugh	$\surd$

Sample VocalSound Instances

Freq Rank	Token Count	VocalSound Description	Used
1	11515	laugh	√
2	7091	breath	
3	4589	inbreath	
4	2223	mouth	
5	970	breath-laugh	√
11	97	laugh-breath	√
46	6	cough-laugh	√
63	3	laugh, "hmmph"	√
69	3	breath while smiling	
75	2	very long laugh	√

- laughter is by far the most common non-verbal VocalSound
- idem for Comment instances

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` laughs

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` laughs
  - 11845 were adjacent to a time-stamped utterance boundary or lexical item: endpoints were derived automatically
  - 725 needed to be segmented manually

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` laughs
  - 11845 were adjacent to a time-stamped utterance boundary or lexical item: endpoints were derived automatically
  - 725 needed to be segmented manually
- found 1108 non-farfield `Comment` laughs
  - all needed to be segmented manually

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` laughs
  - 11845 were adjacent to a time-stamped utterance boundary or lexical item: endpoints were derived automatically
  - 725 needed to be segmented manually
- found 1108 non-farfield `Comment` laughs
  - all needed to be segmented manually
- manual segmententation performed by one annotator, checked by at least one other annotator

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` laughs
  - 11845 were adjacent to a time-stamped utterance boundary or lexical item: endpoints were derived automatically
  - 725 needed to be segmented manually
- found 1108 non-farfield `Comment` laughs
  - all needed to be segmented manually
- manual segmententation performed by one annotator, checked by at least one other annotator
- merging immediately adjacent `VocalSound` and `Comment` instances, and removing transcribed instances for which we found counterevidence, resulted in **13259 bouts**

Speech vs Laughter by Time

- 13259 laugh bouts

Speech vs Laughter by Time

- 13259 laugh bouts
- 110790 talk spurts

Speech vs Laughter by Time

- 13259 laugh bouts
- 110790 talk spurts
- by personal time:

Speech vs Laughter by Time

- 13259 laugh bouts
- 110790 talk spurts
- by personal time:
  - 442.6 hours total recorded audio

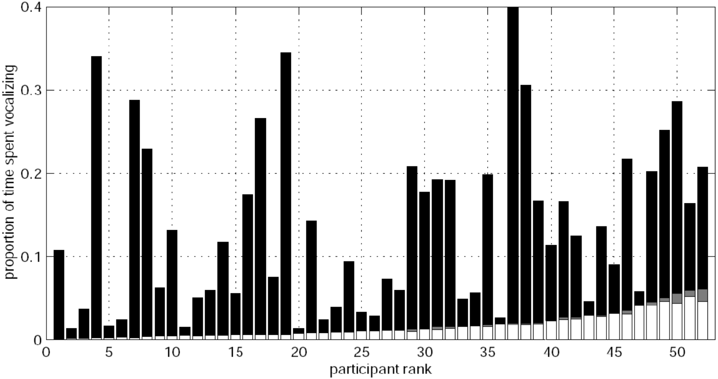
Speech vs Laughter by Time

- 13259 laugh bouts
- 110790 talk spurts
- by personal time:
  - 442.6 hours total recorded audio
  - 55.2 hours spent in talk spurts ($\mathcal{S}$), $\equiv$ 12.47%

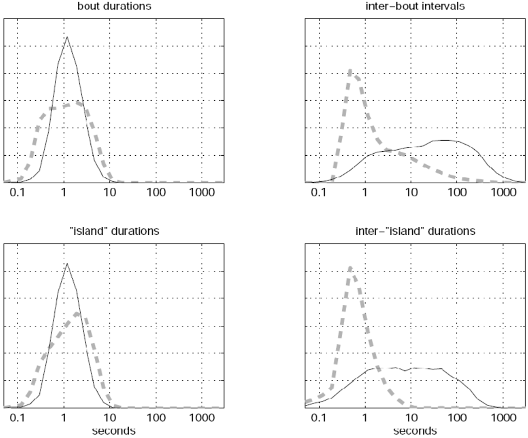
Speech vs Laughter by Time

- 13259 laugh bouts
- 110790 talk spurts
- by personal time:
  - 442.6 hours total recorded audio
  - 55.2 hours spent in talk spurts ($\mathcal{S}$), $\equiv$ 12.47%
  - 5.6 hours spent in laugh bouts ($\mathcal{L}$), $\equiv$ 1.27%

Speech vs Laughter by Time, **by Participant**



Talk Spurt Duration vs Laugh Bout Duration



Vocalization Overlap

Vocal Activity	Vocalizing Time, hrs					
	per part	per meet	number of simultaneously vocalizing participants 1	2	3	$\geq$4
$\mathcal{S}$	55.2	50.8	46.7	3.8	0.27	0.02
$\mathcal{L}$	5.6	3.3	2.0	0.7	0.31	0.27
$\mathcal{S} \cap \mathcal{L}$	0.2	0.2	0.2	0.0	0.0	0
$\mathcal{S} \cup \mathcal{L}$	60.3	52.0	45.7	4.8	0.88	0.49

Vocalization Overlap

Vocal Activity	Vocalizing Time, hrs					
	per part	per meet	number of simultaneously vocalizing participants			
			1	2	3	≥ 4
$\mathcal{S}$	55.2	50.8	46.7	3.8	0.27	0.02
$\mathcal{L}$	5.6	3.3	2.0	0.7	0.31	0.27
$\mathcal{S} \cap \mathcal{L}$	0.2	0.2	0.2	0.0	0.0	0
$\mathcal{S} \cup \mathcal{L}$	60.3	52.0	45.7	4.8	0.88	0.49

- in $\mathcal{S}$ only, 84.6% of vocalization is not overlapped

Vocalization Overlap

Vocal Activity	Vocalizing Time, hrs per part	per meet	number of simultaneously vocalizing participants 1	2	3	≥4
$\mathcal{S}$	55.2	50.8	46.7	3.8	0.27	0.02
$\mathcal{L}$	5.6	3.3	2.0	0.7	0.31	0.27
$\mathcal{S} \cap \mathcal{L}$	0.2	0.2	0.2	0.0	0.0	0
$\mathcal{S} \cup \mathcal{L}$	60.3	52.0	45.7	4.8	0.88	0.49

- in $\mathcal{L}$ only, 35.7% of vocalization is not overlapped

Vocalization Overlap

Vocal Activity	Vocalizing Time, hrs					
	per part	per meet	number of simultaneously vocalizing participants			
			1	2	3	$\geq$4
$\mathcal{S}$	55.2	50.8	46.7	3.8	0.27	0.02
$\mathcal{L}$	5.6	3.3	2.0	0.7	0.31	0.27
$\mathcal{S} \cap \mathcal{L}$	0.2	0.2	0.2	0.0	0.0	0
$\mathcal{S} \cup \mathcal{L}$	60.3	52.0	45.7	4.8	0.88	0.49

- the proportion of “laughed speech” is negligible

Vocalization Overlap

Vocal Activity	Vocalizing Time, hrs					
	per part	per meet	number of simultaneously vocalizing participants			
			1	2	3	$\geq$4
$\mathcal{S}$	55.2	50.8	46.7	3.8	0.27	0.02
$\mathcal{L}$	5.6	3.3	2.0	0.7	0.31	0.27
$\mathcal{S} \cap \mathcal{L}$	0.2	0.2	0.2	0.0	0.0	0
$\mathcal{S} \cup \mathcal{L}$	60.3	52.0	45.7	4.8	0.88	0.49

- there is $\geq$3 times as much 3-participant overlap when considering $\mathcal{S} \cup \mathcal{L}$ as opposed to $\mathcal{S}$ only

Vocalization Overlap

Vocal Activity	Vocalizing Time, hrs					
	per part	per meet	number of simultaneously vocalizing participants			
			1	2	3	$\geq$4
$\mathcal{S}$	55.2	50.8	46.7	3.8	0.27	0.02
$\mathcal{L}$	5.6	3.3	2.0	0.7	0.31	0.27
$\mathcal{S} \cap \mathcal{L}$	0.2	0.2	0.2	0.0	0.0	0
$\mathcal{S} \cup \mathcal{L}$	60.3	52.0	45.7	4.8	0.88	0.49

- there is $\approx$25 times as much 4-participant overlap when considering $\mathcal{S} \cup \mathcal{L}$ as opposed to $\mathcal{S}$ only

Overlap Dynamics

- does laughter differ from speech in the way in which overlap arises and is resolved?

Overlap Dynamics

- does laughter differ from speech in the way in which overlap arises and is resolved?
- look at transition probabilities under a first-order Markov assumption

Overlap Dynamics

- does laughter differ from speech in the way in which overlap arises and is resolved?
- look at transition probabilities under a first-order Markov assumption
    1. discretize $\mathcal{L}$ and $\mathcal{S}$ segmentations using non-overlapping analysis frames

Overlap Dynamics

- does laughter differ from speech in the way in which overlap arises and is resolved?
- look at transition probabilities under a first-order Markov assumption
  1. discretize $\mathcal{L}$ and $\mathcal{S}$ segmentations using non-overlapping analysis frames
  2. train an Extended Degree-of-Overlap (EDO) model on the discretized $\mathcal{L}$ and $\mathcal{S}$ segmentations
     - $P(\{A\} \to \{A, B\})$
     - $P(\{A, B\} \to \{A\})$
     - $P(\{A\} \to \{B\})$
     - etc.

Overlap Dynamics

- does laughter differ from speech in the way in which overlap arises and is resolved?
- look at transition probabilities under a first-order Markov assumption
  1. discretize $\mathcal{L}$ and $\mathcal{S}$ segmentations using non-overlapping analysis frames
  2. train an Extended Degree-of-Overlap (EDO) model on the discretized $\mathcal{L}$ and $\mathcal{S}$ segmentations
     - $P(\{A\} \rightarrow \{A, B\})$
     - $P(\{A, B\} \rightarrow \{A\})$
     - $P(\{A\} \rightarrow \{B\})$
     - etc.
  3. compare inferred probabilities for $\mathcal{L}$ and $\mathcal{S}$

Overlap Dynamics: Results

Select EDO Transitions from (at t)		to (at $t+1$)	500ms frames $\mathcal{S}$	$\mathcal{L}$
$\{A\}$	$\rightarrow$	$\{A\}$	82.94	57.96
$\{A\}$	$\rightarrow$	$\{A, B\}$	6.21	8.43
$\{A\}$	$\rightarrow$	$\{A, B, C, \cdots\}$	0.39	2.39
$\{A, B\}$	$\rightarrow$	$\{A\}$	45.49	26.37
$\{A, B\}$	$\rightarrow$	$\{A, B\}$	40.88	46.93
$\{A, B\}$	$\rightarrow$	$\{A, B, C, \cdots\}$	4.46	13.65
$\{A, B, C, \cdots\}$	$\rightarrow$	$\{A\}$	19.24	6.69
$\{A, B, C, \cdots\}$	$\rightarrow$	$\{A, B\}$	40.94	17.45
$\{A, B, C, \cdots\}$	$\rightarrow$	$\{A, B, C, \cdots\}$	29.44	71.04

Overlap Dynamics: Results

Select EDO Transitions from (at t)		to (at $t+1$)	500ms frames $\mathcal{S}$	$\mathcal{L}$
$\{A\}$	$\rightarrow$	$\{A\}$	82.94	57.96
$\{A\}$	$\rightarrow$	$\{A, B\}$	6.21	8.43
$\{A\}$	$\rightarrow$	$\{A, B, C, \cdots\}$	0.39	2.39
$\{A, B\}$	$\rightarrow$	$\{A\}$	45.49	26.37
$\{A, B\}$	$\rightarrow$	$\{A, B\}$	40.88	46.93
$\{A, B\}$	$\rightarrow$	$\{A, B, C, \cdots\}$	4.46	13.65
$\{A, B, C, \cdots\}$	$\rightarrow$	$\{A\}$	19.24	6.69
$\{A, B, C, \cdots\}$	$\rightarrow$	$\{A, B\}$	40.94	17.45
$\{A, B, C, \cdots\}$	$\rightarrow$	$\{A, B, C, \cdots\}$	29.44	71.04

Conclusions

Based on the ICSI meetings,

1. approximately 9% of vocalizing time is spent on laughter

Conclusions

Based on the ICSI meetings,

1. approximately 9% of vocalizing time is spent on laughter
   - but participants vary widely (0% - 30%)

Conclusions

Based on the ICSI meetings,

1. approximately 9% of vocalizing time is spent on laughter
   - but participants vary widely (0% - 30%)
2. on average, laughter occurs once a minute

Conclusions

Based on the ICSI meetings,

1. approximately 9% of vocalizing time is spent on laughter
   - but participants vary widely (0% - 30%)
2. on average, laughter occurs once a minute
3. laughter accounts for the large majority of $\geq$3 participant overlap

Conclusions

Based on the ICSI meetings,

1. approximately 9% of vocalizing time is spent on laughter
   - but participants vary widely (0% - 30%)
2. on average, laughter occurs once a minute
3. laughter accounts for the large majority of $\geq$3 participant overlap
4. in contrast to speech, once laughter overlap is incurred, it is most likely to persist

Conclusions

Based on the ICSI meetings,

1. approximately 9% of vocalizing time is spent on laughter
   - but participants vary widely (0% - 30%)
2. on average, laughter occurs once a minute
3. laughter accounts for the large majority of $\geq$3 participant overlap
4. in contrast to speech, once laughter overlap is incurred, it is most likely to persist
   - ie. 3-participant speech overlap is 2.5 times more likely than laughter to be resolved within 500 ms

We would like to thank:

- our annotators: Jörg Brunstein and Matthew Bell
- discussion: Alan Black and Liz Shriberg
- funding: EU CHIL