On the Correlation between Perceptual and Contextual Aspects of Laughter in Meetings

Kornel Laskowski & Susanne Burger

interACT, Carnegie Mellon University

August 9, 2007

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

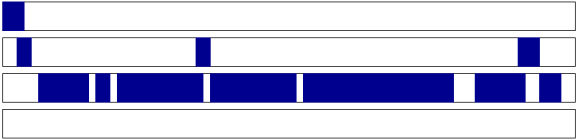
- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

Introduction

- what we do:
  - data-driven, language-/text- independent modeling of
  - multi-participant conversation for
  - automatic conversation recognition and understanding
- why?
  - who has the floor when? how many floors are there?
  - who backchannels when? and towards whom?
  - who interrupts who? who asks questions? who gives answers?
  - how formal is the conversation?
  - what is the social hierarchy of the participants?
  - how do participants appear to feel?

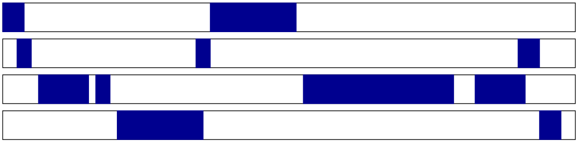
A Text-Independent Representation of Multi-Participant Conversation

- essentially monologue
- "multi-logue"
- heated "multi-logue"
- a mathematical artifact (the Haar wavelet basis)
- "multi-logue" with laughter



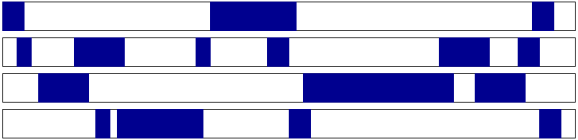
A Text-Independent Representation of Multi-Participant Conversation

- essentially monologue
- "multi-logue"
- heated "multi-logue"
- a mathematical artifact (the Haar wavelet basis)
- "multi-logue" with laughter



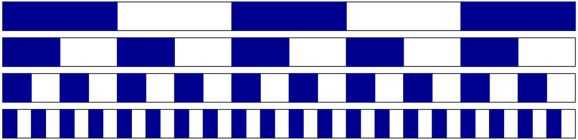
A Text-Independent Representation of Multi-Participant Conversation

- essentially monologue
- "multi-logue"
- heated "multi-logue"
- a mathematical artifact (the Haar wavelet basis)
- "multi-logue" with laughter



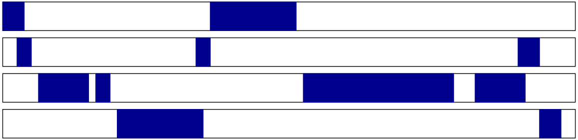
A Text-Independent Representation of Multi-Participant Conversation

- essentially monologue
- "multi-logue"
- heated "multi-logue"
- a mathematical artifact (the Haar wavelet basis)
- "multi-logue" with laughter
  - participants tend to wait their turn to speak
  - participants do not wait to laugh



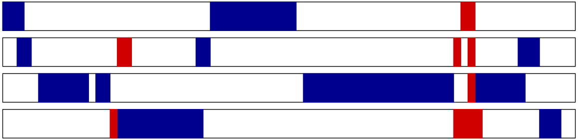
A Text-Independent Representation of Multi-Participant Conversation

- essentially monologue
- “multi-logue”
- heated “multi-logue”
- a mathematical artifact (the Haar wavelet basis)
- “multi-logue” with laughter
  - participants tend to wait their turn to speak
  - participants do not wait to laugh



A Text-Independent Representation of Multi-Participant Conversation

- essentially monologue
- “multi-logue”
- heated “multi-logue”
- a mathematical artifact (the Haar wavelet basis)
- “multi-logue” with laughter
  - participants tend to wait their turn to speak
  - participants **do not** wait to laugh



Emotion and Laughter in Conversation

- external observers of conversation appear to agree as to whether participants feel
  - neutral: 82% of utterances
  - positive: 16% of utterances
  - negative: 2% of utterances
- transcribed laughter is strongly predictive of positive valence (92% classification accuracy)
- A FUTURE GOAL: to find laughter in continuous audio
- context does discriminate between speech and laughter
- **does context discriminate between voiced and unvoiced laughter?**

Emotion and Laughter in Conversation

- external observers of conversation appear to agree as to whether participants feel
  - neutral: 82% of utterances
  - positive: 16% of utterances
  - negative: 2% of utterances
- transcribed laughter is strongly predictive of positive valence (92% classification accuracy)
- A FUTURE GOAL: to find laughter in continuous audio
  - acoustic features
  - context states
- context does discriminate between speech and laughter
- **does context discriminate between voiced and unvoiced laughter?**

Emotion and Laughter in Conversation

- external observers of conversation appear to agree as to whether participants feel
  - neutral: 82% of utterances
  - positive: 16% of utterances
  - negative: 2% of utterances
- transcribed laughter is strongly predictive of positive valence (92% classification accuracy)
- A FUTURE GOAL: to find laughter in continuous audio
  - acoustic features
  - context states
- context does discriminate between speech and laughter
- **does context discriminate between voiced and unvoiced laughter?**

Emotion and Laughter in Conversation

- external observers of conversation appear to agree as to whether participants feel
  - neutral: 82% of utterances
  - positive: 16% of utterances
  - negative: 2% of utterances
- transcribed laughter is strongly predictive of positive valence (92% classification accuracy)
- A FUTURE GOAL: to find laughter in continuous audio
  - acoustic features
  - context states
- context does discriminate between speech and laughter
- **does context discriminate between voiced and unvoiced laughter?**

Emotion and Laughter in Conversation

- external observers of conversation appear to agree as to whether participants feel
  - neutral: 82% of utterances
  - positive: 16% of utterances
  - negative: 2% of utterances
- transcribed laughter is strongly predictive of positive valence (92% classification accuracy)
- A FUTURE GOAL: to find laughter in continuous audio
  - acoustic features
  - context states
- context does discriminate between speech and laughter
- does context discriminate between voiced and unvoiced laughter?

Emotion and Laughter in Conversation

- external observers of conversation appear to agree as to whether participants feel
  - neutral: 82% of utterances
  - positive: 16% of utterances
  - negative: 2% of utterances
- transcribed laughter is strongly predictive of positive valence (92% classification accuracy)
- A FUTURE GOAL: to find laughter in continuous audio
  - acoustic features
  - context states
- context does discriminate between speech and laughter
- **does context discriminate between voiced and unvoiced laughter?**

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations
- for our purposes, 4 types of meetings:

type	# of meetings	# of possible participants	# of participants		
			mod	min	max
Bed	15	13	6	4	7
Bmr	29	15	7	3	9
Bro	23	10	6	4	8
other	8	27	6	5	8

- "other" contains types of which there are $\leq$3 meetings
- types represent longitudinal recordings
- rarely, meetings contain additional, uninstrumented participants

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations
- for our purposes, 4 types of meetings:

type	# of meetings	# of possible participants	# of participants		
			mod	min	max
Bed	15	13	6	4	7
Bmr	29	15	7	3	9
Bro	23	10	6	4	8
other	8	27	6	5	8

- "other" contains types of which there are ≤ 3 meetings
- types represent longitudinal recordings
- rarely, meetings contain additional, uninstrumented participants

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations
- for our purposes, 4 types of meetings:

type	# of meetings	# of possible participants	# of participants mod	# of participants min	# of participants max
Bed	15	13	6	4	7
Bmr	29	15	7	3	9
Bro	23	10	6	4	8
other	8	27	6	5	8

- "other" contains types of which there are $\leq$3 meetings
- types represent longitudinal recordings
- rarely, meetings contain additional, uninstrumented participants

The ICSI Meeting Corpus

- naturally occurring project-oriented conversations
- for our purposes, 4 types of meetings:

type	# of meetings	# of possible participants	# of participants		
			mod	min	max
Bed	15	13	6	4	7
Bmr	29	15	7	3	9
Bro	23	10	6	4	8
other	8	27	6	5	8

- "other" contains types of which there are $\leq$3 meetings
- types represent longitudinal recordings
- rarely, meetings contain additional, uninstrumented participants

The ICSI Meeting Corpus

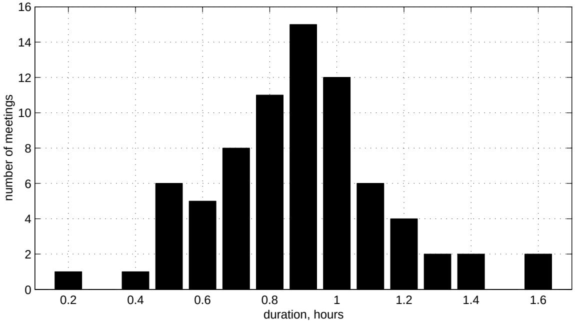
- naturally occurring project-oriented conversations
- for our purposes, 4 types of meetings:

type	# of meetings	# of possible participants	# of participants		
			mod	min	max
Bed	15	13	6	4	7
Bmr	29	15	7	3	9
Bro	23	10	6	4	8
other	8	27	6	5	8

- "other" contains types of which there are $\leq$3 meetings
- types represent longitudinal recordings
- rarely, meetings contain additional, uninstrumented participants

The ICSI Meeting Corpus: Amount of Audio

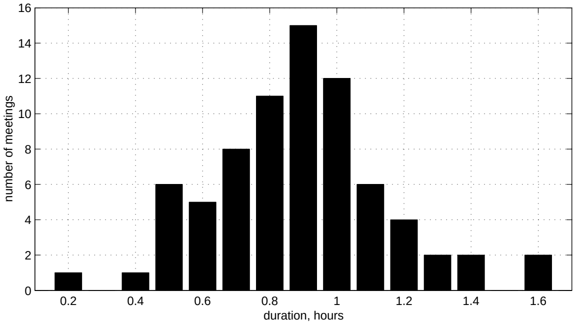
- distribution of usable meeting durations over the 75 meetings:



- a total of 66.3 hours of conversation
- the average participant vocalizes for 14.8% of the time

The ICSI Meeting Corpus: Amount of Audio

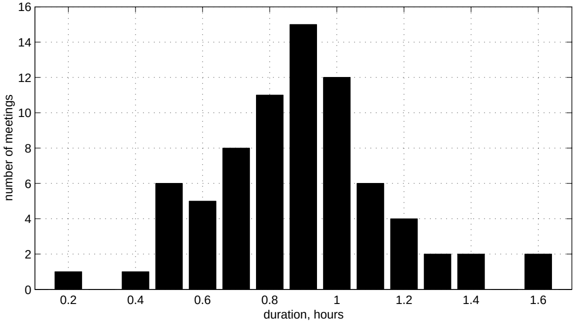
- distribution of usable meeting durations over the 75 meetings:



- a total of 66.3 hours of conversation
- the average participant vocalizes for 14.8% of the time

The ICSI Meeting Corpus: Amount of Audio

- distribution of usable meeting durations over the 75 meetings:



- a total of 66.3 hours of conversation
- the average participant vocalizes for 14.8% of the time

Laughter Annotation

- the ICSI corpus (audio) is accompanied by orthographic transcription, which includes a relatively rich XML-style mark-up of laughter
- for our purposes, data preprocessing consisted of:
  1. identifying laughter in the orthographic transcription
  2. segmentation: specifying endpoints for identified laughter
  3. classification: specifying voicing for segmented laughter

Laughter Annotation

- the ICSI corpus (audio) is accompanied by orthographic transcription, which includes a relatively rich XML-style mark-up of laughter
- for our purposes, data preprocessing consisted of:
  1. identifying laughter in the orthographic transcription
  2. segmentation: specifying endpoints for identified laughter
  3. classification: specifying voicing for segmented laughter

Laughter Annotation

- the ICSI corpus (audio) is accompanied by orthographic transcription, which includes a relatively rich XML-style mark-up of laughter
- for our purposes, data preprocessing consisted of:
  1. identifying laughter in the orthographic transcription
  2. segmentation: specifying endpoints for identified laughter
  3. classification: specifying voicing for segmented laughter

Laughter Annotation

- the ICSI corpus (audio) is accompanied by orthographic transcription, which includes a relatively rich XML-style mark-up of laughter
- for our purposes, data preprocessing consisted of:
  1. identifying laughter in the orthographic transcription
  2. segmentation: specifying endpoints for identified laughter
  3. classification: specifying voicing for segmented laughter

Identifying Laughter in the ICSI Corpus

- 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"/>
```

- laughter is identified using `VocalSound` and `Comment` tags

Identifying Laughter in the ICSI Corpus

- 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"/>
```

- laughter is identified using `VocalSound` and `Comment` tags

Identifying Laughter in the ICSI Corpus

- 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"/>
```

- laughter is identified using `VocalSound` and `Comment` tags

Identifying Laughter in the ICSI Corpus

- 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"/>
```

- laughter is identified using VocalSound and Comment tags

Identifying Laughter in the ICSI Corpus

- 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"/>
```

- laughter is identified using `VocalSound` and `Comment` tags

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 vocal sound annotated in this corpus

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 vocal sound annotated in this corpus

Sample Comment Instances

Freq Rank	Token Count	Comment Description
2	980	while laughing
16	59	while smiling
44	13	last two words while laughing
125	4	last word while laughing
145	3	vocal gesture, a mock laugh

- the most frequent Comment is not related to conversation
- therefore, while laughing is the most frequent conversation-related Comment description
- Comment tags have an even richer description set than VocalSound tags

Sample Comment Instances

Freq Rank	Token Count	Comment Description
2	980	while laughing
16	59	while smiling
44	13	last two words while laughing
125	4	last word while laughing
145	3	vocal gesture, a mock laugh

- the most frequent Comment is not related to conversation
- therefore, while laughing is the most frequent conversation-related Comment description
- Comment tags have an even richer description set than VocalSound tags

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` instances
  - 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` instances
  - all needed to be segmented manually
- manual segmententation performed by me, 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 segmented bouts of laughter

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` instances
  - 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` instances
  - all needed to be segmented manually
- manual segmententation performed by me, 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 segmented bouts of laughter

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` instances
  - 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` instances
  - all needed to be segmented manually
- manual segmententation performed by me, 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 segmented bouts of laughter

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` instances
  - 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` instances
  - all needed to be segmented manually
- manual segmententation performed by me, 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 segmented bouts of laughter

Segmenting Identified Laughter Instances

- found 12570 non-farfield `VocalSound` instances
  - 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` instances
  - all needed to be segmented manually
- manual segmententation performed by me, 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 segmented bouts of laughter

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
  - not modified: [illegible]
  - modified voicing: [illegible]
  - modified endpoints: [illegible]
  - removed: 50 bouts ([illegible])
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
  - not modified: 11961 bouts (90.2%)
  - modified voicing: 942 bouts (7.1%)
  - modified endpoints: 306 bouts (2.3%)
  - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *kappa* was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
  - not modified: 11961 bouts (90.2%)
  - modified voicing: 942 bouts (7.1%)
  - modified endpoints: 306 bouts (2.3%)
  - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
  - not modified: 11961 bouts (90.2%)
  - modified voicing: 942 bouts (7.1%)
  - modified endpoints: 306 bouts (2.3%)
  - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
    - not modified: 11961 bouts (90.2%)
    - modified voicing: 942 bouts (7.1%)
    - modified endpoints: 306 bouts (2.3%)
    - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *kappa* was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
  - not modified: 11961 bouts (90.2%)
  - modified voicing: 942 bouts (7.1%)
  - modified endpoints: 306 bouts (2.3%)
  - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
    - not modified: 11961 bouts (90.2%)
    - modified voicing: 942 bouts (7.1%)
    - modified endpoints: 306 bouts (2.3%)
    - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Classifying Voicing of the Segmented Laughter Bouts

- if any portion of the bout is voiced, the bout is voiced
- performed manually for all 13259 bouts by at least one annotator
- interlabeler *k*appa was 0.76-0.79 (we considered this low)
- all instances rechecked by Susi
  - not modified: 11961 bouts (90.2%)
  - modified voicing: 942 bouts (7.1%)
  - modified endpoints: 306 bouts (2.3%)
  - removed: 50 bouts (0.4%)
- total left: 13209 bouts

Voiced vs Unvoiced Laughter by Time

- of 13209 bouts of laughter,
  - voiced: 8687 (65.8%)
  - unvoiced: 4426 (33.5%)
  - *laughed speech*: 96 (0.7%)
- of 5.7 hours of laughter
  - voiced: 4.2 hours (73.7%)
  - unvoiced: 1.5 hours (25.8%)
  - *laughed speech*: <0.1 hours (0.5%)
- since there is so little *laughed speech*, we ignore it in this work

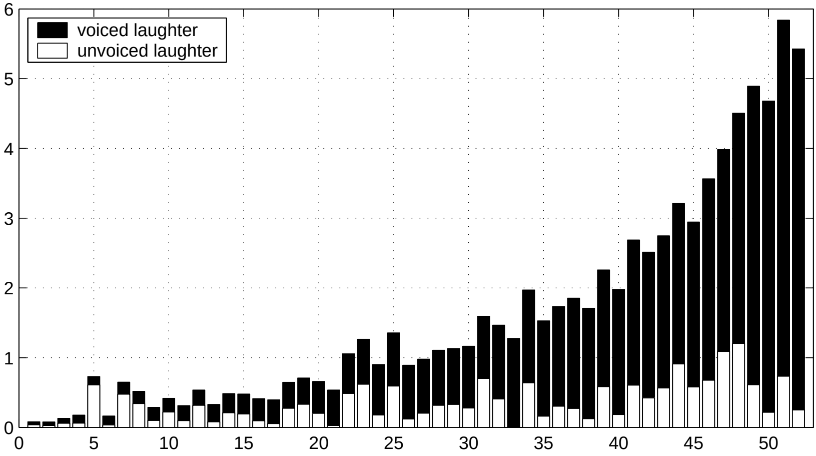
Voiced vs Unvoiced Laughter by Time

- of 13209 bouts of laughter,
  - voiced: 8687 (65.8%)
  - unvoiced: 4426 (33.5%)
  - *laughed speech*: 96 (0.7%)
- of 5.7 hours of laughter
  - voiced: 4.2 hours (73.7%)
  - unvoiced: 1.5 hours (25.8%)
  - *laughed speech*: $<$0.1 hours (0.5%)
- since there is so little *laughed speech*, we ignore it in this work

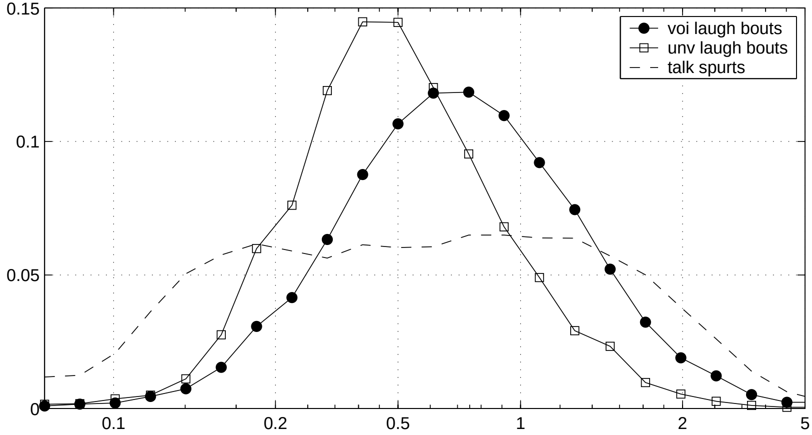
Voiced vs Unvoiced Laughter by Time

- of 13209 bouts of laughter,
  - voiced: 8687 (65.8%)
  - unvoiced: 4426 (33.5%)
  - *laughed speech*: 96 (0.7%)
- of 5.7 hours of laughter
  - voiced: 4.2 hours (73.7%)
  - unvoiced: 1.5 hours (25.8%)
  - *laughed speech*: $<$0.1 hours (0.5%)
- since there is so little *laughed speech*, we ignore it in this work

Voiced vs Unvoiced Laughter by Time, **by Participant**



Voiced vs Unvoiced Bout Duration



Analysis of **Laughter-in-Interaction**

- GOAL: characterize the correlation between voicing in laughter and the vocal interaction context in which laughter occurs
  - test for the statistical significance of association
  - test for the strength of association (predictability)

1. discretize (in time) the voiced laughter, unvoiced laughter, and talkspurt segmentations
2. for each discrete laugh frame, extract a set of multi-participant, participant-independent features from the discretized context
3. characterize the association between context features and voicing features

Analysis of **Laughter-in-Interaction**

- GOAL: characterize the correlation between voicing in laughter and the vocal interaction context in which laughter occurs
  - test for the statistical significance of association
  - test for the strength of association (predictability)

1. discretize (in time) the voiced laughter, unvoiced laughter, and talkspurt segmentations
   - allows for counting
2. for each discrete laugh frame, extract a set of multi-participant, participant-independent features from the discretized context
3. characterize the association between context features and voicing features

Analysis of **Laughter-in-Interaction**

- GOAL: characterize the correlation between voicing in laughter and the vocal interaction context in which laughter occurs
  - test for the statistical significance of association
  - test for the strength of association (predictability)

1. discretize (in time) the voiced laughter, unvoiced laughter, and talkspurt segmentations
   - allows for counting
2. for each discrete laugh frame, extract a set of multi-participant, participant-independent features from the discretized context
3. characterize the association between context features and voicing features

Analysis of **Laughter-in-Interaction**

- GOAL: characterize the correlation between voicing in laughter and the vocal interaction context in which laughter occurs
  - test for the statistical significance of association
  - test for the strength of association (predictability)

1. discretize (in time) the voiced laughter, unvoiced laughter, and talkspurt segmentations
   - allows for counting
2. for each discrete laugh frame, extract a set of multi-participant, participant-independent features from the discretized context
3. characterize the association between context features and voicing features

Analysis of **Laughter-in-Interaction**

- GOAL: characterize the correlation between voicing in laughter and the vocal interaction context in which laughter occurs
  - test for the statistical significance of association
  - test for the strength of association (predictability)

1. discretize (in time) the voiced laughter, unvoiced laughter, and talkspurt segmentations
   - allows for counting
2. for each discrete laugh frame, extract a set of multi-participant, participant-independent features from the discretized context
3. characterize the association between context features and voicing features

Analysis of **Laughter-in-Interaction**

- GOAL: characterize the correlation between voicing in laughter and the vocal interaction context in which laughter occurs
  - test for the statistical significance of association
  - test for the strength of association (predictability)

1. discretize (in time) the voiced laughter, unvoiced laughter, and talkspurt segmentations
   - allows for counting
2. for each discrete laugh frame, extract a set of multi-participant, participant-independent features from the discretized context
3. characterize the association between context features and voicing features

Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:

Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:

Discretizing Segmentations

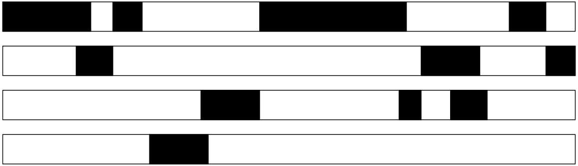
- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:

Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:

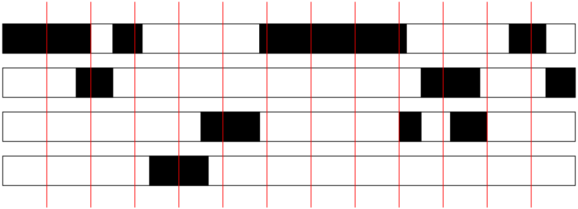
Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



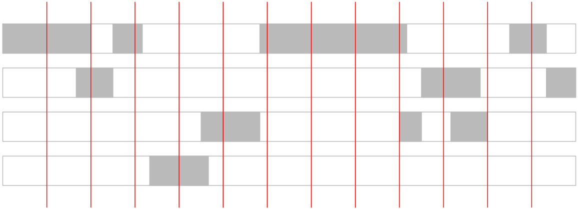
Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



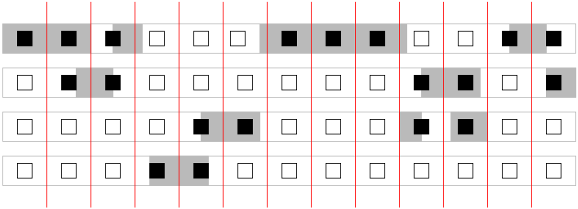
Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



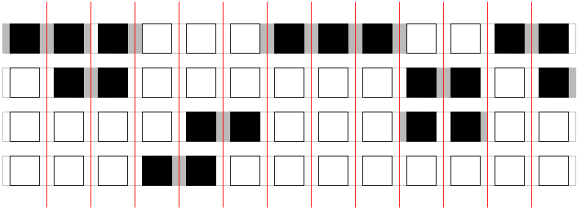
Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



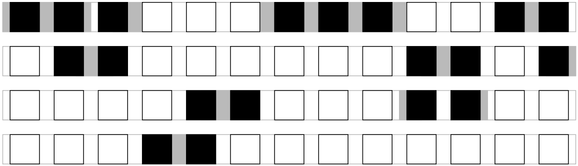
Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



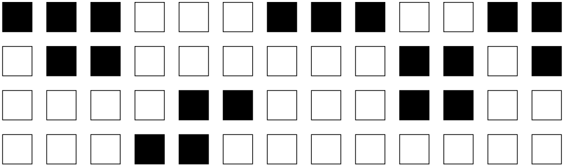
Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



Discretizing Segmentations

- chop up each segmentation into non-overlapping 1 second frames
- for each participant k, declare a frame centered on time t as "on" when participant k vocalizes for at least 10% of that frame's duration
- example:



Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Features Describing Conversational Context

- for each frame t in which participant k laughs:
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a talk spurt
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which contains voicing
  - count how many other participants, at times $t-1$, t, and $t+1$, are producing a laugh bout which does not contain voicing
  - determine whether participant k is speaking at times $t-1$ and $t+1$
- in total, each frame of voiced or unvoiced laughter corresponds to a vocal interaction context defined by 11 features

Summary of Context and Voicing Features

- at this point, have:

	# other participants in									participant k in speech?		Voicing?
	speech			voiced laughter			unvoiced laughter					
	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	$t+1$	
1	1	1	0	0	1	2	0	0	0	N	N	Y
2	0	0	1	0	0	1	0	1	1	Y	N	Y
3	0	1	1	0	2	3	1	0	0	N	Y	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- now, can proceed to analysis

Summary of Context and Voicing Features

- at this point, have:

context features

	# other participants in									participant k in speech?		Voicing?
	speech			voiced laughter			unvoiced laughter					
	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	$t+1$	
1	1	1	0	0	1	2	0	0	0	N	N	Y
2	0	0	1	0	0	1	0	1	1	Y	N	Y
3	0	1	1	0	2	3	1	0	0	N	Y	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- now, can proceed to analysis

Summary of Context and Voicing Features

- at this point, have:

context features

	# other participants in									participant k in speech?		Voicing?
	speech			voiced laughter			unvoiced laughter					
	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	$t+1$	
1	1	1	0	0	1	2	0	0	0	N	N	Y
2	0	0	1	0	0	1	0	1	1	Y	N	Y
3	0	1	1	0	2	3	1	0	0	N	Y	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

integer features

binary features

- now, can proceed to analysis

Summary of Context and Voicing Features

- at this point, have:

context features

	# other participants in									participant k in speech?		Voicing?
	speech			voiced laughter			unvoiced laughter					
	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	t	$t+1$	$t-1$	$t+1$	
1	1	1	0	0	1	2	0	0	0	N	N	Y
2	0	0	1	0	0	1	0	1	1	Y	N	Y
3	0	1	1	0	2	3	1	0	0	N	Y	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

integer features binary features

- now, can proceed to analysis

Testing Significance and Strength of Association

- GOAL: correlate context features with the single voicing feature
- OPTION 1: standard, **one-feature**-at-a-time:
  1. significance: a 2×2 χ^2-test
  2. strength: mutual information (or other entropy-related)
- OPTION 2: optimal ordering of **multiple-features**-at-once:
- latter is known as C4.5; developed for the inference of decision tree classifiers from data

Testing Significance and Strength of Association

- GOAL: correlate context features with the single voicing feature
- OPTION 1: standard, **one-feature**-at-a-time:
  1. significance: a 2×2 χ^2-test
  2. strength: mutual information (or other entropy-related)
- OPTION 2: optimal ordering of **multiple-features**-at-once:
  1. strength: incremental, top-down mutual information
  2. significance: bottom-up χ^2-based pruning
- latter is known as C4.5; developed for the inference of decision tree classifiers from data

Testing Significance and Strength of Association

- GOAL: correlate context features with the single voicing feature
- OPTION 1: standard, **one-feature**-at-a-time:
  1. significance: a 2×2 χ^2-test
  2. strength: mutual information (or other entropy-related)
- OPTION 2: optimal ordering of **multiple-features**-at-once:
  1. strength: incremental, top-down mutual information
  2. significance: bottom-up χ^2-based pruning
- latter is known as C4.5; developed for the inference of decision tree classifiers from data

Testing Significance and Strength of Association

- GOAL: correlate context features with the single voicing feature
- OPTION 1: standard, **one-feature**-at-a-time:
  1. significance: a 2×2 χ^2-test
  2. strength: mutual information (or other entropy-related)
- OPTION 2: optimal ordering of **multiple-features**-at-once:
  1. strength: incremental, top-down mutual information
  2. significance: bottom-up χ^2-based pruning
- latter is known as C4.5; developed for the inference of decision tree classifiers from data

Testing Significance and Strength of Association

- GOAL: correlate context features with the single voicing feature
- OPTION 1: standard, **one-feature**-at-a-time:
  1. significance: a 2×2 χ^2-test
  2. strength: mutual information (or other entropy-related)
- OPTION 2: optimal ordering of **multiple-features**-at-once:
  1. strength: incremental, top-down mutual information
  2. significance: bottom-up χ^2-based pruning
- latter is known as C4.5; developed for the inference of decision tree classifiers from data

Testing Significance and Strength of Association

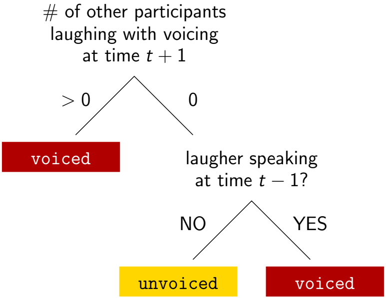
- GOAL: correlate context features with the single voicing feature
- OPTION 1: standard, **one-feature**-at-a-time:
  1. significance: a 2×2 χ^2-test
  2. strength: mutual information (or other entropy-related)
- OPTION 2: optimal ordering of **multiple-features**-at-once:
  1. strength: incremental, top-down mutual information
  2. significance: bottom-up χ^2-based pruning
- latter is known as C4.5; developed for the inference of decision tree classifiers from data

Inferred Decision Tree for Laughter Initiation

- **initiation of laughter**: look at those laughter frames which are the *first* frames of each bout
- the inferred decision tree, χ^2-pruned ($p < 0.05$) to retain only statistically significant nodes:

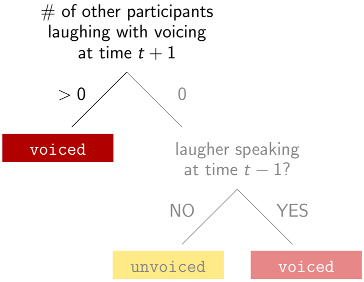
Inferred Decision Tree for Laughter Initiation

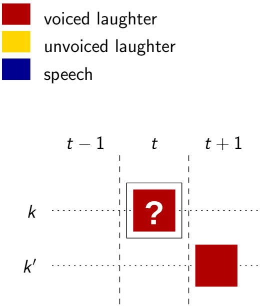
- **initiation of laughter**: look at those laughter frames which are the *first* frames of each bout
- the inferred decision tree, χ^2-pruned ($p < 0.05$) to retain only statistically significant nodes:



Understanding the Laughter Initiation Decision Tree

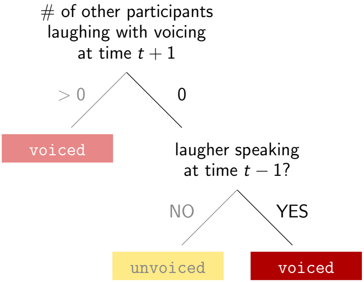
Case 1 when at least one other participant laughs with voicing just after
$\longrightarrow$ **voiced**

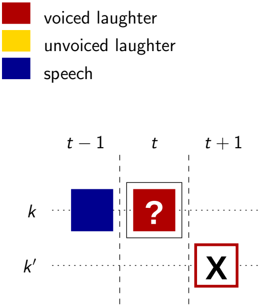




Understanding the Laughter Initiation Decision Tree

Case 2 when no other participants laugh with voicing just after **AND** the laugher speaks just before $\longrightarrow$ **voiced**

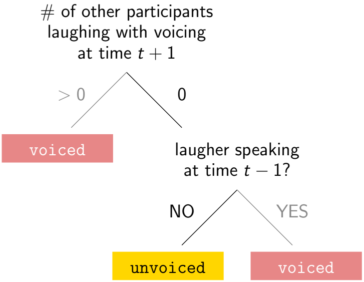


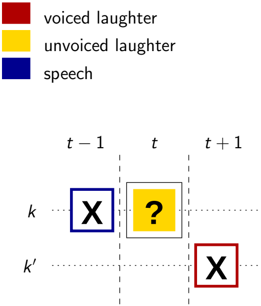


Understanding the Laughter Initiation Decision Tree

Case 3 when no other participants laugh with voicing just after **AND** the laugher does not speak just before

$\longrightarrow$ **unvoiced**



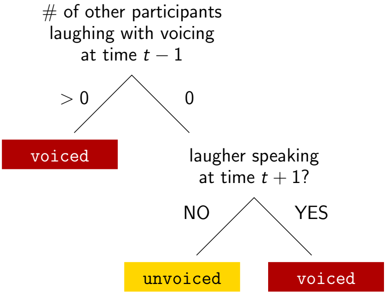


Inferred Decision Tree for Laughter Termination

- **termination of laughter**: look at those laughter frames which are the *last* frames of each bout
- the inferred decision tree, χ^2-pruned ($p < 0.05$) to retain only statistically significant nodes:

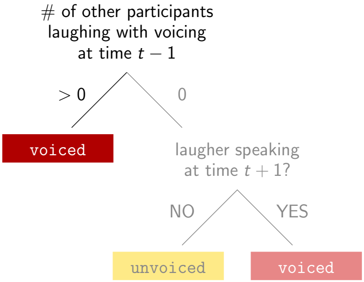
Inferred Decision Tree for Laughter Termination

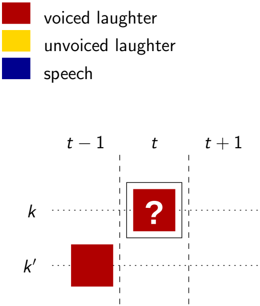
- **termination of laughter**: look at those laughter frames which are the *last* frames of each bout
- the inferred decision tree, χ^2-pruned ($p < 0.05$) to retain only statistically significant nodes:



Understanding the Laughter Termination Decision Tree

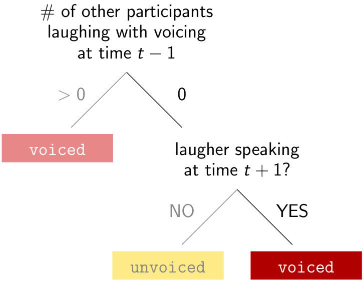
Case 1 when at least one other participant laughs with voicing just before
$\longrightarrow$ **voiced**

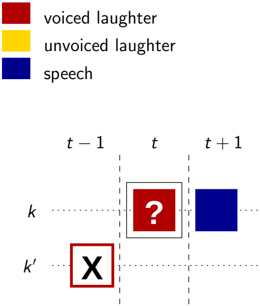




Understanding the Laughter Termination Decision Tree

Case 2 when no other participants laugh with voicing just before **AND** the laugher speaks just after $\longrightarrow$ **voiced**

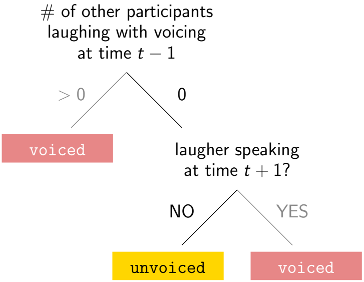


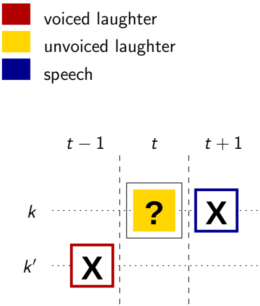


Understanding the Laughter Termination Decision Tree

Case 3 when no other participants laugh with voicing just before **AND** the laugher does not speak just after

$\longrightarrow$ **unvoiced**



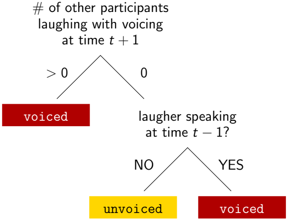


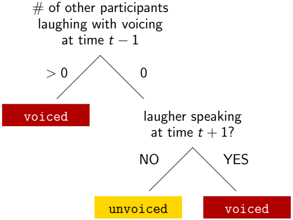
Some Interesting Observations

- we found no statistically significant tree for laughter frames that were neither the first nor the last frame of a bout
- the initiation and termination tree are exactly symmetrical

Some Interesting Observations

- we found no statistically significant tree for laughter frames that were neither the first nor the last frame of a bout
- the initiation and termination tree are exactly symmetrical





Conclusions I

- of 13209 studied bouts of laughter, 66.5% appear to be voiced and 33.5% appear to be unvoiced
- on average, each participant spends approximately 10% of their vocalization effort on laughter (as opposed to speech)
- bout durations follow a log-normal distribution, as expected
  - the mode of voiced laugh bout durations is approximately twice as large as that of unvoiced laugh bout durations
  - but bout duration does not discriminate between voiced and unvoiced laughter

Conclusions I

- of 13209 studied bouts of laughter, 66.5% appear to be voiced and 33.5% appear to be unvoiced
- on average, each participant spends approximately 10% of their vocalization effort on laughter (as opposed to speech)
- bout durations follow a log-normal distribution, as expected
  - the mode of voiced laugh bout durations is approximately twice as large as that of unvoiced laugh bout durations
  - but bout duration does not discriminate between voiced and unvoiced laughter

Conclusions I

- of 13209 studied bouts of laughter, 66.5% appear to be voiced and 33.5% appear to be unvoiced
- on average, each participant spends approximately 10% of their vocalization effort on laughter (as opposed to speech)
- bout durations follow a log-normal distribution, as expected
  - the mode of voiced laugh bout durations is approximately twice as large as that of unvoiced laugh bout durations
  - but bout duration does *not* discrimitate between voiced and unvoiced laughter

Conclusions I

- of 13209 studied bouts of laughter, 66.5% appear to be voiced and 33.5% appear to be unvoiced
- on average, each participant spends approximately 10% of their vocalization effort on laughter (as opposed to speech)
- bout durations follow a log-normal distribution, as expected
  - the mode of voiced laugh bout durations is approximately twice as large as that of unvoiced laugh bout durations
  - but bout duration does *not* discrimitate between voiced and unvoiced laughter

Conclusions I

- of 13209 studied bouts of laughter, 66.5% appear to be voiced and 33.5% appear to be unvoiced
- on average, each participant spends approximately 10% of their vocalization effort on laughter (as opposed to speech)
- bout durations follow a log-normal distribution, as expected
  - the mode of voiced laugh bout durations is approximately twice as large as that of unvoiced laugh bout durations
  - but bout duration does *not* discrimitate between voiced and unvoiced laughter

Conclusions II

1. laughter which begins just before others laugh with voicing and laughter which ends just after others laugh with voicing is likely to be **voiced**
2. when not (1), laughter which begins after the laugher speaks and laughter which ends before the laugher speaks is likely to be **voiced**
3. when not (1) or (2), laughter is likely to be **unvoiced**

Conclusions II

1. laughter which begins just before others laugh with voicing and laughter which ends just after others laugh with voicing is likely to be **voiced**
2. when not (1), laughter which begins after the laugher speaks and laughter which ends before the laugher speaks is likely to be **voiced**
3. when not (1) or (2), laughter is likely to be **unvoiced**

Conclusions II

1. laughter which begins just before others laugh with voicing and laughter which ends just after others laugh with voicing is likely to be **voiced**
2. when not (1), laughter which begins after the laugher speaks and laughter which ends before the laugher speaks is likely to be **voiced**
3. when not (1) or (2), laughter is likely to be **unvoiced**