Introduction	Data 00	Model 000000	Experiments 0000000	Analysis 00000000	Summary

Kornel Laskowski and Tanja Schultz

Cognitive Systems Lab, Universität Karlsruhe Language Technologies Institute, Carnegie Mellon University

09 September, 2008

Introduction ●○○○○○	Data 00	Model 000000	Experiments 0000000	Analysis 000000000	Summary			
Why Detec	Why Detect Laughter in Meetings?							

- evidence suggests that it is the most frequently occurring and most robust behavior which external observers associate with perceived emotion
 - marked valence
 - marked activation
- automatic emotion recognition in meetings

 enable indexing, search and summary, mediated by
 para-propositional content

 also necessary for automorphics mechanis participation

回 と く ヨ と く ヨ と

Introduction ●○○○○○○	Data 00	Model 000000	Experiments 0000000	Analysis 000000000	Summary			
Why Detec	Why Detect Laughter in Meetings?							

- evidence suggests that it is the most frequently occurring and most robust behavior which external observers associate with perceived emotion
 - marked valence
 - marked activation

automatic emotion recognition in meetings

- enable indexing, search and summary, mediated by para-propositional content
- also necessary for autonomous machine participation

O detection and tracking of humour and seriousness

・ 同 ト ・ ヨ ト ・ ヨ ト

Introduction ●○○○○○○	Data 00	Model 000000	Experiments 0000000	Analysis 000000000	Summary			
Why Detec	Why Detect Laughter in Meetings?							

- evidence suggests that it is the most frequently occurring and most robust behavior which external observers associate with perceived emotion
 - marked valence
 - marked activation
- automatic emotion recognition in meetings
 - enable indexing, search and summary, mediated by para-propositional content
 - also necessary for autonomous machine participation
- Output detection and tracking of humour and seriousness

向下 イヨト イヨト

Introduction ●○○○○○○	Data ○○	Model 000000	Experiments	Analysis 00000000	Summary			
Why Detec	Why Detect Laughter in Meetings?							

- evidence suggests that it is the most frequently occurring and most robust behavior which external observers associate with perceived emotion
 - marked valence
 - marked activation
- automatic emotion recognition in meetings
 - enable indexing, search and summary, mediated by para-propositional content
 - also necessary for autonomous machine participation

2 detection and tracking of humour and seriousness

Introduction ●○○○○○○	Data ○○	Model 000000	Experiments	Analysis 00000000	Summary			
Why Detec	Why Detect Laughter in Meetings?							

- evidence suggests that it is the most frequently occurring and most robust behavior which external observers associate with perceived emotion
 - marked valence
 - marked activation
- automatic emotion recognition in meetings
 - enable indexing, search and summary, mediated by para-propositional content
 - also necessary for autonomous machine participation
- 2 detection and tracking of humour and seriousness

Introduction ○●○○○○○	Data oo	Model 000000	Experiments	Analysis 00000000	Summary			
Classifying Emotional Valence								

- data: ISL Meeting Corpus (Burger et al, 2002)
- annotation: perceived valence (Laskowski & Burger, 2006)
- task: classify segmented utteraces as exhibiting one of {negative,neutral,positive}

Classification	Accuracy, %
Classification	Eval
guessing with uniform prior	33.3
guessing with $\mathrm{TrainSet}$ prior	\approx 67
guessing majority $\mathrm{TRAINSET}$ class	\approx 81
presence of $\mathcal L$ only	≈ 92
prosody features	≈84
all features (except presence of $\mathcal{L})$	\approx 87

Introduction ○○●○○○○	Data 00	Model 000000	Experiments	Analysis 000000000	Summary

Classifying Emotional Activation

- also known as emotional arousal
- data: ICSI Meeting Corpus (Janin et al, 2003)
- annotation: hotspots (Wrede & Shriberg, 2004; Wrede et al, 2005)
- task: classify 60-second intervals as one of {involvementContaining, ¬involvementContaining}

Classification	Accuracy, %			
Classification	TRAIN	Dev	EVAL	
guessing with uniform prior	50.0	50.0	50.0	
guessing with TRAINSET prior	61.3	60.9	61.2	
guessing majority $\operatorname{TRAINSET}$ class	73.7	72.9	73.7	
features from ${\cal L}$ only	79.2	80.0	80.6	
features from ${\cal L}$ and ${\cal S}$	84.3	82.7	83.0	

Introduction	Data oo	Model	Experiments	Analysis	Summary
Goals of t	this Work	Ţ.			

- propose a framework for detecting laughter from audio only
 - close-talk microphones on all participants
- attempt to detect all laughter
 - temporally isolated from the laugher's speech
 - occurring within dialog acts among verbal productions
- attempt to detect without prior knowledge
 - no inactive channel exclusion
 - expect to encounter many false alarms
- attribute laughter to specific participants

Introduction ○○○●○○○	Data oo	Model	Experiments	Analysis	Summary
Goals of	this Work				

propose a framework for detecting laughter from audio only
 close-talk microphones on all participants

attempt to detect all laughter

temporally isolated from the laugher's speech

- occurring within dialog acts among verbal productions
- attempt to detect without prior knowledge
 - no inactive channel exclusion
 - expect to encounter many false alarms
- attribute laughter to specific participants

・ 同 ト ・ ヨ ト ・ ヨ ト

Introduction	Data oo	Model	Experiments	Analysis 00000000	Summary
Goals of t	this Work				

- propose a framework for detecting laughter from audio only
 - close-talk microphones on all participants
- attempt to detect all laughter
 - temporally isolated from the laugher's speech
 - **2** occurring within dialog acts among verbal productions
- attempt to detect without prior knowledge
 - no inactive channel exclusion
 - expect to encounter many false alarms
- attribute laughter to specific participants

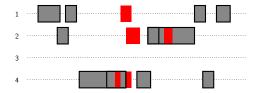
Introduction	Data 00	Model	Experiments	Analysis	Summary
Goals of t	his Work				

- propose a framework for detecting laughter from audio only
 - close-talk microphones on all participants
- attempt to detect all laughter
 - temporally isolated from the laugher's speech
 - **2** occurring within dialog acts among verbal productions
- 3 attempt to detect without prior knowledge
 - no inactive channel exclusion
 - expect to encounter many false alarms
 - attribute laughter to specific participants

Introduction	Data oo	Model 000000	Experiments	Analysis 000000000	Summary
Goals of t	his Work				

- propose a framework for detecting laughter from audio only
 - close-talk microphones on all participants
- attempt to detect all laughter
 - temporally isolated from the laugher's speech
 - occurring within dialog acts among verbal productions
- attempt to detect without prior knowledge
 - no inactive channel exclusion
 - expect to encounter many false alarms
- attribute laughter to specific participants

Introduction	Data 00	Model 000000	Experiments	Analysis 00000000	Summary

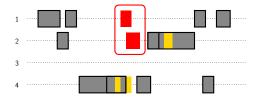


Past work has focused on:

- a subset of laughter (improving recall)
 - isolated laughter
 - loud, clear, unambiguous laughter
- and/or a subset of audio (improving precision)

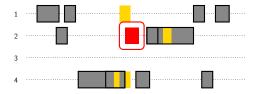
- 4 同下 4 ヨト 4 ヨト

Introduction	Data 00	Model 000000	Experiments	Analysis 00000000	Summary



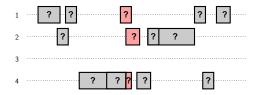
- a subset of laughter (improving recall)
 - isolated laughter
 - loud, clear, unambiguous laughter
- and/or a subset of audio (improving precision)
 - segmented intervals
 - only active channels

Introduction	Data 00	Model 000000	Experiments	Analysis 00000000	Summary



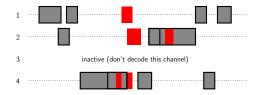
- a subset of laughter (improving recall)
 - isolated laughter
 - loud, clear, unambiguous laughter
- and/or a subset of audio (improving precision)
 segmented intervals
 a collumetries chappels

Introduction	Data 00	Model 000000	Experiments	Analysis	Summary



- a subset of laughter (improving recall)
 - isolated laughter
 - loud, clear, unambiguous laughter
- and/or a subset of audio (improving precision)
 - segmented intervals
 - only active channels

Introduction ○○○○●○○	Data oo	Model 000000	Experiments	Analysis 00000000	Summary



- a subset of laughter (improving recall)
 - isolated laughter
 - loud, clear, unambiguous laughter
- and/or a subset of audio (improving precision)
 - segmented intervals
 - only active channels

Introduction ○○○○○●○	Data 00	Model 000000	Experiments	Analysis 000000000	Summary

Brief Comparison with Related Work

Aspect	\mathcal{L}/\mathcal{S}	class.	$\mathcal{L}/$ -	$\neg \mathcal{L}$ se	gm.	this
Aspect	[1]	[2]	[3]	[4]	[5]	work
close-talk microphones	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
farfield microphones					\checkmark	
single channel at-a-time	\checkmark	\checkmark	\checkmark	\checkmark		
multi-channel at-a-time					\checkmark	\checkmark
participant attribution	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
only group laughter					\checkmark	
only isolated laughter	\checkmark		\checkmark	\checkmark		
only clear laughter		\checkmark				
rely on pre-segmentation	\checkmark	\checkmark	?			
rely on channel exclusion			?	\checkmark		

[1] (Truong & van Leeuwen, 2005); [2] (Truong & van Leeuwen, 2007a); [3] (Truong & van Leeuwen, 2007b); [4] (Knox & Mirghafori, 2007); [5] (Kennedy & Ellis, 2004).

Introduction	Data	Model	Experiments	Analysis	Summary
○○○○○○●	oo	000000	0000000	000000000	000
Outline c	of this Tal	k			

- 1. Introduction (about to be over)
- 2. Data
- 3. Multiparticipant 3-state Vocal Activity Detector
- 4. Experiments
- 5. Analysis
- 6. Conclusions (& Unqualified Recommendations)

Introduction	Data ●○	Model 000000	Experiments 0000000	Analysis 00000000	Summary
ICSI Mee	ting Corp	ous			

• the complete corpus (Janin et al, 2003)

- 75 naturally occurring meetings
- longitudinal CTM recordings of several work groups
- 3-9 instrumented participants per meeting

we use a subset of 67 meetings

- types: Bed (15), Bmr (29), Bro (23)
- 23 unique participants
- 3 participants attend both Bmr and Bro
- 1 participant attends both Bmr and Bed
- in particular, as elsewhere,
 - TRAINSET: 26 Bmr meetings
 - TESTSET: 3 Bmr meetings

・ 同 ト ・ ヨ ト ・ ヨ

Introduction	Data ●○	Model 000000	Experiments	Analysis 000000000	Summary
ICSI Mee	ting Corp	ous			

- the complete corpus (Janin et al, 2003)
 - 75 naturally occurring meetings
 - longitudinal CTM recordings of several work groups
 - 3-9 instrumented participants per meeting
- we use a subset of 67 meetings
 - types: Bed (15), Bmr (29), Bro (23)
 - 23 unique participants
 - 3 participants attend both Bmr and Bro
 - 1 participant attends both Bmr and Bed
- in particular, as elsewhere,
 - TRAINSET: 26 Bmr meetings
 - TESTSET: 3 Bmr meetings

Introduction	Data ●○	Model 000000	Experiments	Analysis 000000000	Summary
ICSI Mee	ting Corp	ous			

- the complete corpus (Janin et al, 2003)
 - 75 naturally occurring meetings
 - longitudinal CTM recordings of several work groups
 - 3-9 instrumented participants per meeting
- we use a subset of 67 meetings
 - types: Bed (15), Bmr (29), Bro (23)
 - 23 unique participants
 - 3 participants attend both Bmr and Bro
 - 1 participant attends both Bmr and Bed
- in particular, as elsewhere,
 - TRAINSET: 26 Bmr meetings
 - TESTSET: 3 Bmr meetings

Introduction	Data ○●	Model 000000	Experiments	Analysis	Summary
	~				

Reference Segmentation

• speech, \mathcal{S}

- forced alignment of words and word fragments
- available in the ICSI MRDA Corpus (Shriberg et al, 2004)
- bridge inter-lexeme gaps shorter than 300 ms
- as in NIST Rich Transcription Meeting Recognition evaluations
- \bullet laughter, $\mathcal L$
 - produced semi-automatically (Laskowski & Burger, 2007d)
 - $\bullet~\geq$ 99% of laughter markup, as originally transcribed
 - bouts include terminal "recovery" in-/ehxalation, if present
 - \bullet augmented with voicing classification, $\mathcal{L} \equiv \mathcal{L}_V \cup \mathcal{L}_U$
- "laughed speech" (Nwokah et al, 1999), $\mathcal{S}\cap\mathcal{L}$
 - $\bullet\,$ here, mapped to laughter ${\cal L}$
 - each participant can be producing \mathcal{L} , \mathcal{S} , or neither

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	○●	000000	0000000	000000000	000
Reference	Segmen	tation			

- ${\scriptstyle \bullet}$ speech, ${\cal S}$
 - forced alignment of words and word fragments
 - available in the ICSI MRDA Corpus (Shriberg et al, 2004)
 - bridge inter-lexeme gaps shorter than 300 ms
 - as in NIST Rich Transcription Meeting Recognition evaluations
- laughter, $\mathcal L$
 - produced semi-automatically (Laskowski & Burger, 2007d)
 - $\bullet~\geq$ 99% of laughter markup, as originally transcribed
 - bouts include terminal "recovery" in-/ehxalation, if present
 - \bullet augmented with voicing classification, $\mathcal{L} \equiv \mathcal{L}_V \cup \mathcal{L}_U$
- "laughed speech" (Nwokah et al, 1999), $\mathcal{S}\cap\mathcal{L}$
 - $\bullet\,$ here, mapped to laughter ${\cal L}$
 - each participant can be producing \mathcal{L} , \mathcal{S} , or neither

(4月) (1日) (日)

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	○●	000000		000000000	000
Reference	- Segmen	tation			

- ${\scriptstyle \bullet}$ speech, ${\cal S}$
 - forced alignment of words and word fragments
 - available in the ICSI MRDA Corpus (Shriberg et al, 2004)
 - bridge inter-lexeme gaps shorter than 300 ms
 - as in NIST Rich Transcription Meeting Recognition evaluations
- laughter, $\mathcal L$
 - produced semi-automatically (Laskowski & Burger, 2007d)
 - $\bullet~\geq$ 99% of laughter markup, as originally transcribed
 - bouts include terminal "recovery" in-/ehxalation, if present
 - \bullet augmented with voicing classification, $\mathcal{L} \equiv \mathcal{L}_V \cup \mathcal{L}_U$
- "laughed speech" (Nwokah et al, 1999), $\mathcal{S}\cap\mathcal{L}$
 - here, mapped to laughter \mathcal{L}
 - $\bullet\,$ each participant can be producing $\mathcal{L},\,\mathcal{S},\,$ or neither

A (1) > A (1) > A (1)

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	○●	000000	0000000	000000000	000
Reference	- Segmen	tation			

- ${\scriptstyle \bullet}$ speech, ${\cal S}$
 - forced alignment of words and word fragments
 - available in the ICSI MRDA Corpus (Shriberg et al, 2004)
 - bridge inter-lexeme gaps shorter than 300 ms
 - as in NIST Rich Transcription Meeting Recognition evaluations
- laughter, $\mathcal L$
 - produced semi-automatically (Laskowski & Burger, 2007d)
 - $\bullet~\geq$ 99% of laughter markup, as originally transcribed
 - bouts include terminal "recovery" in-/ehxalation, if present
 - augmented with voicing classification, $\mathcal{L} \equiv \mathcal{L}_V \cup \mathcal{L}_U$
- ullet "laughed speech" (Nwokah et al, 1999), $\mathcal{S}\cap\mathcal{L}$
 - $\bullet\,$ here, mapped to laughter ${\cal L}$
 - $\bullet\,$ each participant can be producing $\mathcal{L},\,\mathcal{S},\,$ or neither

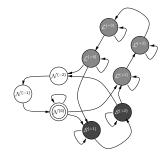
向下 イヨト イヨト

Introduction	Data oo	Model ●○○○○○	Experiments	Analysis 000000000	Summary

Multiparticipant 3-state Vocal Activity Detector

- hidden Markov model
- pruned Viterbi (beam) decoding
- topology
 - single participant state subspace
 - multiparticipant state space, pruning
- multiparticipant transition probability model
- standard MFCC features, plus crosstalk suppression features
- multiparticipant emission probability model

Introduction	Data	Model	Experiments	Analysis	Summary
		000000			



- each participant can be
 - speaking, ${\cal S}$
 - laughing, \mathcal{L}
 - $\bullet\,$ silent, ${\cal N}$

- frame step $\Delta T = 0.1$ s
- explicit minimum duration constraints

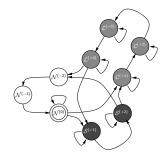
$$\begin{aligned} \mathbf{T}_{min} &\equiv \left(T^{\mathcal{S}}_{min}, T^{\mathcal{L}}_{min}, T^{\mathcal{N}}_{min}\right) \\ &= \Delta T \cdot \left(N^{\mathcal{S}}_{min}, N^{\mathcal{L}}_{min}, N^{\mathcal{N}}_{min}\right) \end{aligned}$$

 number of states in 1-participant subspace

$$N = N_{min}^{S} + N_{min}^{\mathcal{L}} + N_{min}^{\mathcal{N}}$$

- 4 同 ト 4 ヨ ト 4 ヨ

Introduction	Data	Model	Experiments	Analysis	Summary
		000000			



- each participant can be
 - speaking, ${\cal S}$
 - laughing, \mathcal{L}
 - $\bullet\,$ silent, ${\cal N}$

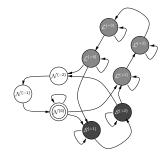
• frame step $\Delta T = 0.1$ s

- explicit minimum duration constraints
 - $\begin{aligned} \mathbf{T}_{min} &\equiv \left(T^{\mathcal{S}}_{min}, T^{\mathcal{L}}_{min}, T^{\mathcal{N}}_{min}\right) \\ &= \Delta T \cdot \left(N^{\mathcal{S}}_{min}, N^{\mathcal{L}}_{min}, N^{\mathcal{N}}_{min}\right) \end{aligned}$
- number of states in 1-participant subspace

$$N = N_{min}^{\mathcal{S}} + N_{min}^{\mathcal{L}} + N_{min}^{\mathcal{N}}$$

(4月) (4日) (4日)

Introduction	Data	Model	Experiments	Analysis	Summary
		00000			



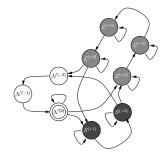
- each participant can be
 - speaking, ${\cal S}$
 - laughing, \mathcal{L}
 - $\bullet\,$ silent, ${\cal N}$

- frame step $\Delta T = 0.1$ s
- explicit minimum duration constraints
 - $$\begin{split} \mathbf{T}_{min} &\equiv \left(T^{\mathcal{S}}_{min}, T^{\mathcal{L}}_{min}, T^{\mathcal{N}}_{min}\right) \\ &= \Delta T \cdot \left(N^{\mathcal{S}}_{min}, N^{\mathcal{L}}_{min}, N^{\mathcal{N}}_{min}\right) \end{split}$$
- number of states in 1-participant subspace

$$N = N_{min}^{\mathcal{S}} + N_{min}^{\mathcal{L}} + N_{min}^{\mathcal{N}}$$

マロト イヨト イヨト

Introduction	Data	Model	Experiments	Analysis	Summary
		00000			



- each participant can be
 - speaking, ${\cal S}$
 - laughing, \mathcal{L}
 - $\bullet\,$ silent, ${\cal N}$

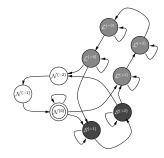
- frame step $\Delta T = 0.1$ s
- explicit minimum duration constraints

$$\begin{aligned} \mathbf{T}_{min} &\equiv \left(T_{min}^{\mathcal{S}}, T_{min}^{\mathcal{L}}, T_{min}^{\mathcal{N}}\right) \\ &= \Delta T \cdot \left(N_{min}^{\mathcal{S}}, N_{min}^{\mathcal{L}}, N_{min}^{\mathcal{N}}\right) \end{aligned}$$

• number of states in 1-participant subspace

$$N = N_{min}^{S} + N_{min}^{\mathcal{L}} + N_{min}^{\mathcal{N}}$$

Introduction	Data	Model	Experiments	Analysis	Summary
		00000			



- each participant can be
 - speaking, ${\cal S}$
 - laughing, \mathcal{L}
 - $\bullet\,$ silent, ${\cal N}$

- frame step $\Delta T = 0.1$ s
- explicit minimum duration constraints

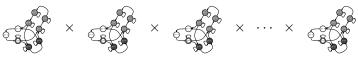
$$\begin{aligned} \mathbf{T}_{min} &\equiv \left(T_{min}^{\mathcal{S}}, T_{min}^{\mathcal{L}}, T_{min}^{\mathcal{N}}\right) \\ &= \Delta T \cdot \left(N_{min}^{\mathcal{S}}, N_{min}^{\mathcal{L}}, N_{min}^{\mathcal{N}}\right) \end{aligned}$$

• number of states in 1-participant subspace

$$N = N_{min}^{\mathcal{S}} + N_{min}^{\mathcal{L}} + N_{min}^{\mathcal{N}}$$

Introduction	Data oo	Model ○○●○○○	Experiments	Analysis 000000000	Summary		
Multiparticipant (MP) State Space							

- for a conversation of K participants,
- form the Cartesian product of K factors:

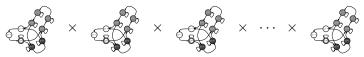


- each MP state: *K*-vector of *N* SP states
- total number of MP states in topology: N¹
- impose maximum simultaneous vocalization constraints

$$\mathbf{K}_{max} = \left(K_{max}^{\mathcal{S}}, K_{max}^{\mathcal{L}}, K_{max}^{\neg \mathcal{N}}\right)$$

Introduction	Data oo	Model ○○●○○○	Experiments 0000000	Analysis 000000000	Summary		
Multiparticipant (MP) State Space							

- for a conversation of K participants,
- form the Cartesian product of K factors:



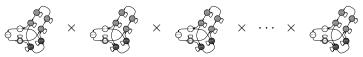
- each MP state: K-vector of N SP states
- total number of MP states in topology: N⁴
- impose maximum simultaneous vocalization constraints

$$\mathbf{K}_{max} = \left(K_{max}^{\mathcal{S}}, K_{max}^{\mathcal{L}}, K_{max}^{\neg \mathcal{N}}\right)$$

• ie. $K_{max}^{\mathcal{L}}$: max. # participants laughing at the same time

Introduction	Data 00	Model ○○●○○○	Experiments	Analysis 00000000	Summary		
Multiparticipant (MP) State Space							

- for a conversation of K participants,
- form the Cartesian product of K factors:

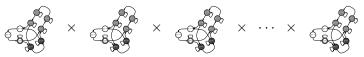


- each MP state: K-vector of N SP states
- total number of MP states in topology: N^{K}
- impose maximum simultaneous vocalization constraints $\mathbf{W} = (\mathbf{W}^{S} + \mathbf{W}^{T} + \mathbf{W}^{T})$

• ie. $K_{max}^{\mathcal{L}}$: max. # participants laughing at the same time

Introduction	Data 00	Model ○○●○○○	Experiments	Analysis 00000000	Summary
Multipart	icipant (I	MP) State	Space		

- for a conversation of K participants,
- form the Cartesian product of K factors:



- each MP state: K-vector of N SP states
- total number of MP states in topology: N^{K}
- impose maximum simultaneous vocalization constraints

$$\mathbf{K}_{max} = \left(K_{max}^{\mathcal{S}}, K_{max}^{\mathcal{L}}, K_{max}^{\neg \mathcal{N}} \right)$$

• ie. $K_{max}^{\mathcal{L}}$: max. # participants laughing at the same time

Introduction	Data 00	Model ○○○●○○	Experiments	Analysis 00000000	Summary

• Example, K = 5:

- at time *t*, $\mathbf{q}_t = \mathbf{S}_i = [S^{(2)}, N^{(0)}, N^{(0)}, N^{(0)}]$
- at time t+1, $\mathbf{q}_{t+1} = \mathbf{S}_j = \left[\mathcal{N}^{(-2)}, \ \mathcal{N}^{(0)}, \ \mathcal{S}^{(1)}, \ \mathcal{L}^{(1)}\right]$
- what is $a_{ij} = P(\mathbf{q}_{t+1} = \mathbf{S}_j | \mathbf{q}_t = \mathbf{S}_i)$?

 a_{ij} = 0 if the SP transition from S_i to S_j for any participant is not licensed by the SP topology

- Otherwise, ML estimate using ngram counts from best flat-start Viterbi path over training corpus
- NOTE: each participant's index k in S is arbitrary
 - for all K-symbol permutations/rotations R
 - want $P(\mathbf{S}_j | \mathbf{S}_i) \equiv P(\mathbf{R} \cdot \mathbf{S}_j | \mathbf{R} \cdot \mathbf{S}_i)$
 - during model training & querying, rotate each q_t into a fixed ordering of the N single-participant states

< ロト < 同ト < ヨト < ヨト

Introduction	Data 00	Model ○○○●○○	Experiments	Analysis 00000000	Summary

• Example, K = 5:

- at time *t*, $\mathbf{q}_t = \mathbf{S}_i = [S^{(2)}, N^{(0)}, N^{(0)}, N^{(0)}]$
- at time t+1, $\mathbf{q}_{t+1} = \mathbf{S}_j = \left[\mathcal{N}^{(-2)}, \ \mathcal{N}^{(0)}, \ \mathcal{S}^{(1)}, \ \mathcal{L}^{(1)}\right]$
- what is $a_{ij} = P(\mathbf{q}_{t+1} = \mathbf{S}_j | \mathbf{q}_t = \mathbf{S}_i)$?
- a_{ij} = 0 if the SP transition from S_i to S_j for any participant is not licensed by the SP topology
- Otherwise, ML estimate using ngram counts from best flat-start Viterbi path over training corpus
- **INOTE:** each participant's index k in **S** is arbitrary
 - for all K-symbol permutations/rotations R
 - want $P(\mathbf{S}_j | \mathbf{S}_i) \equiv P(\mathbf{R} \cdot \mathbf{S}_j | \mathbf{R} \cdot \mathbf{S}_i)$
 - during model training & querying, rotate each q_t into a fixed ordering of the N single-participant states

- 4 同 2 4 三 2 4 三

Introduction	Data 00	Model ○○○●○○	Experiments	Analysis 00000000	Summary

• Example, K = 5:

- at time *t*, $\mathbf{q}_t = \mathbf{S}_i = [S^{(2)}, N^{(0)}, N^{(0)}, N^{(0)}]$
- at time t+1, $\mathbf{q}_{t+1} = \mathbf{S}_j = \left[\mathcal{N}^{(-2)}, \ \mathcal{N}^{(0)}, \ \mathcal{S}^{(1)}, \ \mathcal{L}^{(1)}\right]$
- what is $a_{ij} = P(\mathbf{q}_{t+1} = \mathbf{S}_j | \mathbf{q}_t = \mathbf{S}_i)$?
- a_{ij} = 0 if the SP transition from S_i to S_j for any participant is not licensed by the SP topology
- otherwise, ML estimate using ngram counts from best flat-start Viterbi path over training corpus
- NOTE: each participant's index k in S is arbitrary
 - for all K-symbol permutations/rotations R
 - want $P(\mathbf{S}_i | \mathbf{S}_i) \equiv P(\mathbf{R} \cdot \mathbf{S}_i | \mathbf{R} \cdot \mathbf{S}_i]$
 - during model training & querying, rotate each q_t into a fixed ordering of the N single-participant states

(不同) とうり くうり

Introduction	Data oo	Model ○○○●○○	Experiments	Analysis 00000000	Summary

- Example, K = 5:
 - at time *t*, $\mathbf{q}_t = \mathbf{S}_i = [S^{(2)}, N^{(0)}, N^{(0)}, N^{(0)}]$
 - at time t+1, $\mathbf{q}_{t+1} = \mathbf{S}_j = \left[\mathcal{N}^{(-2)}, \ \mathcal{N}^{(0)}, \ \mathcal{S}^{(1)}, \ \mathcal{L}^{(1)}\right]$
 - what is $a_{ij} = P(\mathbf{q}_{t+1} = \mathbf{S}_j | \mathbf{q}_t = \mathbf{S}_i)$?
- a_{ij} = 0 if the SP transition from S_i to S_j for any participant is not licensed by the SP topology
- otherwise, ML estimate using ngram counts from best flat-start Viterbi path over training corpus
- Solution NOTE: each participant's index k in S is arbitrary
 - for all K-symbol permutations/rotations R
 - want $P(\mathbf{S}_j | \mathbf{S}_i) \equiv P(\mathbf{R} \cdot \mathbf{S}_j | \mathbf{R} \cdot \mathbf{S}_i)$
 - during model training & querying, rotate each q_t into a fixed ordering of the N single-participant states

向下 イヨト イヨト

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	oo	○○○○●○		000000000	000
Observabl	es				

• each of K participants is wearing a close-talk mic (CTM)

• extract 41 features from every CTM channel

- log energy + MFCCs
- Δs and $\Delta \Delta s$
- min and max normalized log energy differences (NLEDs) (Boakye & Stolcke, 2006)
- 41·K features per frame
- may vary from meeting to meeting (as K does)

Introduction	Data oo	Model ○○○●○	Experiments	Analysis 000000000	Summary
Observable	es				

- each of K participants is wearing a close-talk mic (CTM)
- extract 41 features from every CTM channel
 - Iog energy + MFCCs
 - Δs and $\Delta \Delta s$
 - min and max normalized log energy differences (NLEDs) (Boakye & Stolcke, 2006)
- 41·K features per frame
- may vary from meeting to meeting (as K does)

- 4 回 ト 4 ヨ ト 4 ヨ ト

Introduction	Data oo	Model ○○○●○	Experiments	Analysis 000000000	Summary
Observable	es				

- each of K participants is wearing a close-talk mic (CTM)
- extract 41 features from every CTM channel
 - Iog energy + MFCCs
 - Δs and $\Delta \Delta s$
 - min and max normalized log energy differences (NLEDs) (Boakye & Stolcke, 2006)
- 41·K features per frame
- may vary from meeting to meeting (as K does)

向下 イヨト イヨト

Introduction	Data oo	Model ○○○●○	Experiments	Analysis 000000000	Summary
Observable	es				

- each of K participants is wearing a close-talk mic (CTM)
- extract 41 features from every CTM channel
 - Iog energy + MFCCs
 - Δs and $\Delta \Delta s$
 - min and max normalized log energy differences (NLEDs) (Boakye & Stolcke, 2006)
- 41·K features per frame
- may vary from meeting to meeting (as K does)

向下 イヨト イヨト

Introduction	Data oo	Model ○○○○○●	Experiments	Analysis	Summary

Emission Probability Model

• variable feature length vector $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_K]$

train a single-channel GMM (64 components)

- for ${\mathcal S}$ and ${\mathcal L}$
- for \mathcal{N}_{all} and $\mathcal{N}_{nearfield}$
- then approximate the joint MP emission with

$$P(\mathbf{X}|\mathbf{S}_{i}) = \prod_{k=1}^{K} P(\mathbf{X}[k] | \mathbf{S}_{i}[k])$$
(1)

(4月) (4日) (4日)

Introduction	Data oo	Model ○○○○○●	Experiments	Analysis 000000000	Summary
Emission	Probabili	ity Model			

- variable feature length vector $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_K]$
- train a single-channel GMM (64 components)
 - $\bullet~\mbox{for}~{\cal S}~\mbox{and}~{\cal L}$
 - for $\mathcal{N}_{\textit{all}}$ and $\mathcal{N}_{\textit{nearfield}}$

then approximate the joint MP emission with

 $P(\mathbf{X}|\mathbf{S}_i) = \prod_{k=1}^{K} P(\mathbf{X}[k]|\mathbf{S}_i[k])$

白 と く ヨ と く ヨ と

Introduction	Data 00	Model ○○○○●	Experiments	Analysis 000000000	Summary
Emission	Probabili	ity Model			

- variable feature length vector $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_K]$
- train a single-channel GMM (64 components)
 - for ${\mathcal S}$ and ${\mathcal L}$
 - for $\mathcal{N}_{\textit{all}}$ and $\mathcal{N}_{\textit{nearfield}}$
- then approximate the joint MP emission with

$$P(\mathbf{X} | \mathbf{S}_i) = \prod_{k=1}^{K} P(\mathbf{X}[k] | \mathbf{S}_i[k])$$
(1)

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	oo	000000	●○○○○○○	000000000	000
Described	d Experim	nents			

- independent versus joint participant decoding
- sensitivity to minimum duration constraints
- sensitivity to maximum overlap constraints
- generalization to other (non-Bmr) meetings

Evaluation: recall (R), precision (P), and unweighted F

- goal here: \mathcal{L} versus $\neg \mathcal{L} = \mathcal{S} \cup \mathcal{N}$
- sanity: S versus $\neg S = \mathcal{L} \cup \mathcal{N}$
- sanity: $\mathcal{V} = \mathcal{S} \cup \mathcal{L}$ versus $\neg \mathcal{V} = \mathcal{N}$

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	oo	000000	●○○○○○○	000000000	000
Described	d Experim	nents			

- independent versus joint participant decoding
- sensitivity to minimum duration constraints
- sensitivity to maximum overlap constraints
- generalization to other (non-Bmr) meetings

Evaluation: recall (R), precision (P), and unweighted F

- goal here: \mathcal{L} versus $\neg \mathcal{L} = \mathcal{S} \cup \mathcal{N}$
- sanity: S versus $\neg S = \mathcal{L} \cup \mathcal{N}$
- sanity: $\mathcal{V} = \mathcal{S} \cup \mathcal{L}$ versus $\neg \mathcal{V} = \mathcal{N}$

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	oo	000000	●○○○○○○	000000000	000
Described	d Experim	nents			

- independent versus joint participant decoding
- sensitivity to minimum duration constraints
- sensitivity to maximum overlap constraints
- generalization to other (non-Bmr) meetings

Evaluation: recall (R), precision (P), and unweighted F

- goal here: \mathcal{L} versus $\neg \mathcal{L} = \mathcal{S} \cup \mathcal{N}$
- sanity: S versus $\neg S = \mathcal{L} \cup \mathcal{N}$
- sanity: $\mathcal{V} = \mathcal{S} \cup \mathcal{L}$ versus $\neg \mathcal{V} = \mathcal{N}$



Single-participant vs Multiparticipant Decoding

- for decoding participants independently
 - $\mathcal{N}_{\textit{all}}$ and $\mathcal{N}_{\textit{farfield}}$ both represent nearfield silence \mathcal{N}
 - $\bullet~\rightarrow$ 3 competing models, rather than 4
- for decoding participant jointly, can use either 3 or 4 models

Decoding	\mathcal{V}		S			\mathcal{L}	
Decouning	F	R	Р	F	R	Р	F
indep, 3 AM	76.3	90.3	85.0	87.6	80.9	20.4	32.6
joint, 3 AM	78.8	89.7	86.0	87.8	59.2	20.6	30.6
► joint, 4 AM	79.5	83.6	90.0	86.7	55.2	25.1	34.5

- joint decoding improves precision by reducing potential overlap
- 2 modeling farfield vocalization on CTMs significantly improves precision for ${\cal S}$ and ${\cal L}$

イロト イボト イヨト イヨト



Single-participant vs Multiparticipant Decoding

- for decoding participants independently
 - $\mathcal{N}_{\textit{all}}$ and $\mathcal{N}_{\textit{farfield}}$ both represent nearfield silence \mathcal{N}
 - $\bullet~\rightarrow$ 3 competing models, rather than 4
- for decoding participant jointly, can use either 3 or 4 models

Decoding	\mathcal{V}		S			\mathcal{L}	
Decouning	F	R	Р	F	R	Р	F
indep, 3 AM	76.3	90.3	85.0	87.6	80.9	20.4	32.6
joint, 3 AM	78.8	89.7	86.0	87.8	59.2	20.6	30.6
► joint, 4 AM	79.5	83.6	90.0	86.7	55.2	25.1	34.5

- joint decoding improves precision by reducing potential overlap
- 0 modeling farfield vocalization on CTMs significantly improves precision for $\mathcal S$ and $\mathcal L$

イロト イボト イヨト イヨト



Single-participant vs Multiparticipant Decoding

- for decoding participants independently
 - $\mathcal{N}_{\textit{all}}$ and $\mathcal{N}_{\textit{farfield}}$ both represent nearfield silence \mathcal{N}
 - $\bullet~\rightarrow$ 3 competing models, rather than 4
- for decoding participant jointly, can use either 3 or 4 models

Decoding	\mathcal{V}		S			\mathcal{L}	
Decouning	F	R	Р	F	R	Р	F
indep, 3 AM	76.3	90.3	85.0	87.6	80.9	20.4	32.6
joint, <mark>3 AM</mark>	78.8	89.7	86.0	87.8	59.2	20.6	30.6
► joint, <mark>4 AM</mark>	79.5	83.6	90.0	86.7	55.2	25.1	34.5

- joint decoding improves precision by reducing potential overlap
- 0 modeling farfield vocalization on CTMs significantly improves precision for $\mathcal S$ and $\mathcal L$

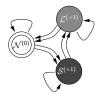
イロト イボト イヨト イヨト

Introduction	Data 00	Model	Experiments	Analysis	Summary
Alternativ	e Minim	um Durati	on Constrair	nts T _{min}	

$\mathbf{T}_{min} = (0.1, 0.1, 0.1) \qquad \mathbf{T}_{min} = (0.3, 0.3, 0.3) \qquad \mathbf{T}_{min} = (0.2, 0.4, 0.3)$

◆□> ◆□> ◆目> ◆目> ◆目> □目 − つへぐ

Introduction	Data 00	Model 000000	Experiments	Analysis 00000000	Summary



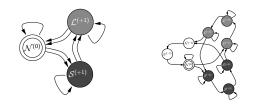
 $\mathbf{T}_{min} = (0.1, 0.1, 0.1) \qquad \mathbf{T}_{min} = (0.3, 0.3, 0.3) \qquad \mathbf{T}_{min} = (0.2, 0.4, 0.3)$

・ 回 ト ・ ヨ ト ・ ヨ ト

 Introduction
 Data
 Model
 Experiments
 Analysis
 Summary

 0000000
 00
 000000
 0000000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000

Alternative Minimum Duration Constraints T_{min}



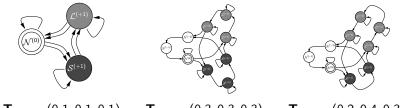
 $\mathbf{T}_{min} = (0.1, 0.1, 0.1)$ $\mathbf{T}_{min} = (0.1, 0.1, 0.1)$

$\mathbf{T}_{min} = (0.3, 0.3, 0.3)$ $\mathbf{T}_{min} = (0.2, 0.4, 0.3)$

- (E

K. Laskowski & T. Schultz Detection of Laughter-in-Interaction in Meetings

Alternative Minimum Duration Constraints T_{min}



 $\mathbf{T}_{min} = (0.1, 0.1, 0.1)$

 $\mathbf{T}_{min} = (0.3, 0.3, 0.3)$

 $\mathbf{T}_{min} = (0.2, 0.4, 0.3)$



• hold maximum overlap constraints fixed, $\mathbf{K}_{max} = (2,3,3)$

T _{min} (s)	\mathcal{V}		S			\mathcal{L}	
• min (3)	F	R	Р	F	R	Р	F
(0.1, 0.1, 0.1)	78.1	82.3	89.9	86.0	55.9		31.7
(0.3, 0.3, 0.3)	79.5	83.7	90.4	86.9	54.7	24.2	33.6
► (0.2, 0.4, 0.3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing all T_{min} from 0.1s to 0.3s improves all F measures
 allowing T^L_{min} > T^S_{min} can result in higher F(L)

イロト イポト イヨト イヨト

-



• hold maximum overlap constraints fixed, $\mathbf{K}_{max} = (2,3,3)$

T _{min} (s)	\mathcal{V}		S			\mathcal{L}	
• min (3)	F	R	Р	F	R	Р	F
(0.1, 0.1, 0.1)	78.1	82.3	89.9	86.0	55.9		31.7
(0.3, 0.3, 0.3)	79.5	83.7	90.4	86.9	54.7	24.2	33.6
▶ (0.2, 0.4, 0.3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing all *T_{min}* from 0.1s to 0.3s improves all F measures
 allowing *T^L_{min} > T^S_{min}* can result in higher F(*L*)

・ロン ・回と ・ヨン・



• hold maximum overlap constraints fixed, $\mathbf{K}_{max} = (2,3,3)$

T _{min} (s)	\mathcal{V}		S			\mathcal{L}	
• min (3)	F	R	Р	F	R	Р	F
(0.1, 0.1, 0.1)	78.1	82.3	89.9	86.0	55.9	22.1	31.7
(0.3, 0.3, 0.3)	79.5	83.7	90.4	86.9	54.7	24.2	33.6
► (0.2, <mark>0.4</mark> , 0.3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

• increasing all T_{min} from 0.1s to 0.3s improves all F measures • allowing $T_{min}^{\mathcal{L}} > T_{min}^{\mathcal{S}}$ can result in higher $F(\mathcal{L})$

・ロン ・回と ・ヨン・

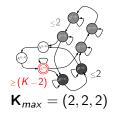
Introduction	Data 00	Model 000000	Experiments	Analysis 000000000	Summary

$K_{max} = (3, 2, 3)$

$\mathbf{K}_{max} = (2, 2, 2)$ $\mathbf{K}_{max} = (2, 2, 3)$

・ロン ・回と ・ヨン ・ヨン

Introduction	Data oo	Model 000000	Experiments	Analysis 000000000	Summary

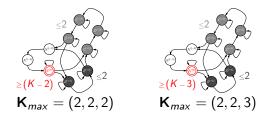


- 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □

 Introduction
 Data
 Model
 Experiments
 Analysis
 Summary

 0000000
 00
 000000
 0000000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000

Alternative Maximum Overlap Constraints K_{max}

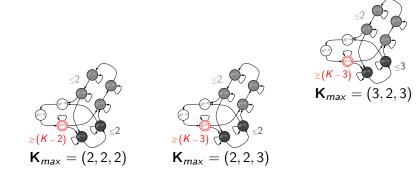


2

<ロ> (四) (四) (三) (三)

IntroductionDataModelExperimentsAnalysisSummary0000000000000000000000000000000

Alternative Maximum Overlap Constraints K_{max}



 $K_{max} = (2, 3, 3)$

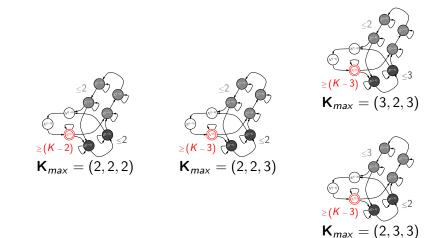
2

(日) (四) (王) (王)

 Introduction
 Data
 Model
 Experiments
 Analysis
 Summary

 0000000
 00
 0000000
 00000000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000

Alternative Maximum Overlap Constraints K_{max}



<ロ> (四) (四) (三) (三)

Introduction	Data 00	Model 000000	Experiments ○○○○●○	Analysis 00000000	Summary

• minimum duration constraints fixed, $T_{min} = (0.2, 0.4, 0.3)$

K _{max}	\mathcal{V}		S			\mathcal{L}	
™ max	F	R	Р	F	R	Р	F
(2,2,2)	81.3	83.3	90.6	86.8	36.9	27.8	31.7
(2, 2, 3)	79.9	84.0	89.0	86.4	48.8	24.3	32.4
(3,2,3)	79.9	84.2	88.6	86.4	49.1	24.6	32.8
► (2, 3, 3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing K_{max} generally leads to higher R and lower P
 increasing K^S_{max} from 2 to 3 has negligible impact
 increasing K^L_{max} from 2 to 3 has significant impact

(4月) (1日) (日)

Introduction	Data 00	Model 000000	Experiments ○○○○●○	Analysis 00000000	Summary

• minimum duration constraints fixed, $T_{min} = (0.2, 0.4, 0.3)$

K _{max}	\mathcal{V}		S			\mathcal{L}	
™ max	F	R	Р	F	R	Р	F
(2,2,2)	81.3	83.3	90.6	86.8	36.9	27.8	31.7
(2,2,3)	79.9	84.0	89.0	86.4	48.8	24.3	32.4
(3, 2, 3)	79.9	84.2	88.6	86.4	49.1	24.6	32.8
► (2, 3, 3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing K_{max} generally leads to higher R and lower P
 increasing K^S_{max} from 2 to 3 has negligible impact
 increasing K^L_{max} from 2 to 3 has significant impact
 because a higher proportion of C is produced in overlap

Introduction	Data 00	Model 000000	Experiments ○○○○○●○	Analysis	Summary

• minimum duration constraints fixed, $T_{min} = (0.2, 0.4, 0.3)$

K _{max}	\mathcal{V}	S			\mathcal{L}		
™ max	F	R	Р	F	R	Р	F
(2,2,2)	81.3	83.3	90.6	86.8	36.9	27.8	31.7
(<mark>2</mark> , 2, 3)	79.9	84.0	89.0	86.4	48.8	24.3	32.4
(3, 2, 3)	79.9	84.2	88.6	86.4	49.1	24.6	32.8
▶ (2,3,3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing K_{max} generally leads to higher R and lower P
 increasing K^S_{max} from 2 to 3 has negligible impact
 increasing K^L_{max} from 2 to 3 has significant impact
 ★ because a higher proportion of L is produced in overlap

伺い イヨト イヨト

Introduction	Data 00	Model 000000	Experiments ○○○○○●○	Analysis	Summary

• minimum duration constraints fixed, $T_{min} = (0.2, 0.4, 0.3)$

K _{max}	\mathcal{V}	S			L		
™ max	F	R	Р	F	R	Р	F
(2,2,2)	81.3	83.3	90.6	86.8	36.9	27.8	31.7
(2, <mark>2</mark> ,3)	79.9	84.0	89.0	86.4	48.8	24.3	32.4
(3,2,3)	79.9	84.2	88.6	86.4	49.1	24.6	32.8
► (2, 3 , 3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing K_{max} generally leads to higher R and lower P
increasing K^S_{max} from 2 to 3 has negligible impact
increasing K^L_{max} from 2 to 3 has significant impact

向下 イヨト イヨ

Introduction	Data oo	Model 000000	Experiments ○○○○○●○	Analysis 00000000	Summary

• minimum duration constraints fixed, $\mathbf{T}_{min} = (0.2, 0.4, 0.3)$

K _{max}	\mathcal{V}		S			\mathcal{L}	
™ max	F	R	Р	F	R	Р	F
(2,2,2)	81.3	83.3	90.6	86.8	36.9	27.8	31.7
(2,2,3)	79.9	84.0	89.0	86.4	48.8	24.3	32.4
(3,2,3)	79.9	84.2	88.6	86.4	49.1	24.6	32.8
► (2, 3, 3)	79.5	83.6	90.0	86.7	55.2	25.1	34.5

increasing K_{max} generally leads to higher R and lower P
 increasing K^S_{max} from 2 to 3 has negligible impact
 increasing K^L_{max} from 2 to 3 has significant impact
 ★ because a higher proportion of L is produced in overlap

• • = • • =

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Generalization to Other Meetings

Test data		$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}	S			\mathcal{L}		
			F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

 ${f D}$ ${f F}({f V})$: ${f Bmr}({f train})>{f Bmr}({f test})>{f Bro}>{f Bed}$

• Bmc(train) and Bmc(test) have lots of participants in common about the Bmc Brochhares & participants, and Bed & participants \odot F(\mathcal{L}) on Bmc(test) higher than on Bmc(train)

<回と < 回と < 回

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Test	data	$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}		S			\mathcal{L}	
TCSU	uata	$\mathcal{P}_{\mathcal{V}}(\mathcal{L})$	F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
DIIII	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

• F(\mathcal{V}): Bmr(train) > Bmr(test) > Bro > Bed

• Bmr(train) and Bmr(test) have lots of participants in common

• with Bmr, Bro shares 3 participants, and Bed 1 participant

P (L) on Bmr(test) higher than on Bmr(train)

・ 同 ト ・ ヨ ト ・ ヨ

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Test	data	$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}		S			\mathcal{L}	
Test	uata	$\mathcal{PV}(\mathcal{L})$	F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
DIIII	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

• F(\mathcal{V}): Bmr(train) > Bmr(test) > Bro > Bed

• Bmr(train) and Bmr(test) have lots of participants in common

• with Bmr, Bro shares 3 participants, and Bed 1 participant

P (L) on Bmr(test) higher than on Bmr(train)

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Test	data	$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}		S			\mathcal{L}	
TCSU	uata	$\mathcal{P}_{\mathcal{V}}(\mathcal{L})$	F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
DIIII	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

• F(\mathcal{V}): Bmr(train) > Bmr(test) > Bro > Bed

• Bmr(train) and Bmr(test) have lots of participants in common

- with Bmr, Bro shares 3 participants, and Bed 1 participant
- **2** $F(\mathcal{L})$ on Bmr(test) higher than on Bmr(train)

 appears to correlate with p_V (L), the proportion of vocalization effort spent on laughter

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Test	data	$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}		S			\mathcal{L}	
Test	uata	$\mathcal{PV}(\mathcal{L})$	F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
DIIIT	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

• F(\mathcal{V}): Bmr(train) > Bmr(test) > Bro > Bed

- Bmr(train) and Bmr(test) have lots of participants in common
- with Bmr, Bro shares 3 participants, and Bed 1 participant
- **2** $F(\mathcal{L})$ on Bmr(test) higher than on Bmr(train)
 - appears to correlate with $p_{\mathcal{V}}(\mathcal{L})$, the proportion of vocalization effort spent on laughter
 - this test set is not typical of the corpus

A (10) A (10) A (10) A

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Test	data	$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}		S			\mathcal{L}	
Test	uata	$\mathcal{PV}(\mathcal{L})$	F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
DIIIT	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

• F(\mathcal{V}): Bmr(train) > Bmr(test) > Bro > Bed

- Bmr(train) and Bmr(test) have lots of participants in common
- with Bmr, Bro shares 3 participants, and Bed 1 participant
- **2** $F(\mathcal{L})$ on Bmr(test) higher than on Bmr(train)
 - appears to correlate with $p_{\mathcal{V}}(\mathcal{L})$, the proportion of vocalization effort spent on laughter
 - this test set is *not typical* of the corpus

- 4 同 ト 4 三 ト 4 三 ト

Introduction	Data	Model	Experiments	Analysis	Summary
			000000		

Test	data	$p_{\mathcal{V}}(\mathcal{L})$	\mathcal{V}		S			\mathcal{L}	
1030	uata	$\mathcal{P}_{\mathcal{V}}(\mathcal{L})$	F	R	Р	F	R	Р	F
Bmr	train	10.91	80.1	83.4	89.8	86.5	53.0	19.4	28.4
DIIIT	test	14.94	79.5	83.6	90.0	86.7	55.2	25.1	34.5
Bro	(all)	5.94	78.1	81.1	90.6	85.6	57.8	11.4	19.0
Bed	(all)	7.53	75.1	84.6	85.7	85.2	58.7	10.0	17.0

• F(\mathcal{V}): Bmr(train) > Bmr(test) > Bro > Bed

- Bmr(train) and Bmr(test) have lots of participants in common
- with Bmr, Bro shares 3 participants, and Bed 1 participant
- **2** $F(\mathcal{L})$ on Bmr(test) higher than on Bmr(train)
 - appears to correlate with $p_{\mathcal{V}}(\mathcal{L})$, the proportion of vocalization effort spent on laughter
 - this test set is *not typical* of the corpus

Introduction	Data 00	Model 000000	Experiments	Analysis ●○○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	2
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}	6.5	9.1	1.0	10.4
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

- final system on test set (13.8 hours)
- all quantities in minutes

글 에 에 글 어

Introduction	Data oo	Model 000000	Experiments	Analysis ○●○○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	2
\mathcal{N}	685.4	22.9	7.8	716.2
L	6.5	9.1	1.0	16.6
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

- break down references $\mathcal{L} \equiv \{ \mathcal{L}'_{U}, \mathcal{L}'_{V}, \mathcal{L} \cap \mathcal{S} \}$
 - L'_U ≡ L_U − L ∩ S: unvoiced laughter less "laughed speech"
 L'_V ≡ L_V − L ∩ S: voiced laughter less "laughed speech"

- 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 回 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □ 2 - 4 □

Introduction	Data oo	Model 000000	Experiments	Analysis ○●○○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	2
\mathcal{N}	685.4	22.9	7.8	716.2
L	6.5	9.1	1.0	16.6
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

- break down references $\mathcal{L} \equiv \{ \mathcal{L}'_{U}, \mathcal{L}'_{V}, \mathcal{L} \cap S \}$
 - L'_U ≡ L_U − L ∩ S: unvoiced laughter less "laughed speech"
 L'_V ≡ L_V − L ∩ S: voiced laughter less "laughed speech"

A (10) A (10) A (10) A

Introduction	Data oo	Model 000000	Experiments	Analysis ○●○○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	2
\mathcal{N}	685.4	22.9	7.8	716.2
L	6.5	9.1	1.0	16.6
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

• break down references $\mathcal{L} \equiv \{ \mathcal{L}'_U, \mathcal{L}'_V, \mathcal{L} \cap S \}$

- $\mathcal{L}'_U \equiv \mathcal{L}_U \mathcal{L} \cap \mathcal{S}$: unvoiced laughter less "laughed speech"
- $\mathcal{L}'_V \equiv \mathcal{L}_V \mathcal{L} \cap \mathcal{S}$: voiced laughter less "laughed speech"

向下 イヨト イヨト

Introduction	Data oo	Model 000000	Experiments 0000000	Analysis ○○●○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	2
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
\mathcal{L}'_V	3.6	6.5	0.3	10.4
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

most unvoiced laughter (L'_U) is classified as silence (N)
 most "laughed speech" (L ∩ S) is classified as speech (S)

・ 回 ・ ・ ヨ ・ ・ ヨ ・

2

Introduction	Data oo	Model 000000	Experiments 0000000	Analysis ○○●○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	2
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
$\mathcal{L}_{V}^{'}$	3.6	6.5	0.3	10.4
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

1 most unvoiced laughter (\mathcal{L}'_U) is classified as silence (\mathcal{N})

② most "laughed speech" $(\mathcal{L}\cap\mathcal{S})$ is classified as speech (\mathcal{S})

ヨト イヨト イヨト

Introduction	Data oo	Model 000000	Experiments	Analysis ○○●○○○○○○	Summary

	hypo	Σ		
	\mathcal{N}	2		
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
$\mathcal{L}_{V}^{'}$	3.6	6.5	0.3	10.4
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

• most unvoiced laughter (\mathcal{L}'_U) is classified as silence (\mathcal{N})

2 most "laughed speech" $(\mathcal{L} \cap \mathcal{S})$ is classified as speech (\mathcal{S})

Introduction	Data 00	Model 000000	Experiments 0000000	Analysis ○○○●○○○○○	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ	
\mathcal{N}	685.4	22.9	7.8	716.2	
\mathcal{L}'_U	2.8	2.4	0.2	5.4	
\mathcal{L}'_V	3.6	6.5	0.3	10.4	\mathcal{L}'
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8	\mathcal{L} (
S	11.0	4.5	79.0	94.4	${\mathcal S}$
Σ	702.9	36.6	87.8	827.2	

	\mathcal{L}	${\mathcal S}$
\mathcal{L}'	8.9	0.5
$\mathcal{L}\cap\mathcal{S}$	0.2	0.5
${\mathcal S}$	4.5	79.0

.≣...>

- - E - F

• looking at \mathcal{L} and \mathcal{S} only

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○●○○○○	Summary

						Recall:
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ		
\mathcal{N}	685.4	22.9	7.8	716.2		
\mathcal{L}'_{U}	2.8	2.4	0.2	5.4		${\cal L}$
$\mathcal{L}_{V}^{'}$	3.6	6.5	0.3	10.4	\mathcal{L}'	94.7
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8	$\mathcal{L}\cap\mathcal{S}$	28.6
${\mathcal S}$	11.0	4.5	79.0	94.4	${\mathcal S}$	5.4
Σ	702.9	36.6	87.8	827.2		

94% of speech is hypothesized as speech

95% of laughter (excluding "laughed speech") is hypothesized as laughter

 \mathcal{S}

5.3

71.4

93.6

2

(4回) (4回) (4回)

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○●○○○○	Summary

						Recall.	
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ		ricean.	
\mathcal{N}	685.4	22.9	7.8	716.2			
\mathcal{L}'_{U}	2.8	2.4	0.2	5.4		\mathcal{L}	
$\mathcal{L}_{V}^{}$	3.6	6.5	0.3	10.4	\mathcal{L}'	94.7	
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8	$\mathcal{L}\cap\mathcal{S}$	28.6	
${\mathcal S}$	11.0	4.5	79.0	94.4	${\mathcal S}$	5.4	
Σ	702.9	36.6	87.8	827.2			
	$\mathcal{L}'_V \ \mathcal{L} \cap \mathcal{S}$	$\begin{array}{ccc} \mathcal{L}'_U & 2.8 \\ \mathcal{L}'_V & 3.6 \\ \mathcal{L} \cap S & 0.1 \\ \mathcal{S} & 11.0 \end{array}$	$\begin{array}{c c} \mathcal{L}'_U & 2.8 & \textbf{2.4} \\ \mathcal{L}'_V & 3.6 & \textbf{6.5} \\ \mathcal{L} \cap \mathcal{S} & 0.1 & \textbf{0.2} \\ \mathcal{S} & 11.0 & 4.5 \end{array}$	$\begin{array}{c} \mathcal{L}'_U \\ \mathcal{L}'_V \\ \mathcal{L}'_V \\ \mathcal{S} \\ \mathcal{S} \\ \mathcal{S} \end{array} \begin{array}{c} 2.8 \\ 2.4 \\ 0.2 \\ 0.2 \\ 0.5 \\ 0.1 \\ 0.2 \\ 0.5 \\ 0.5 \end{array}$	$ \begin{array}{c ccccc} \mathcal{L}'_U & 2.8 & \textbf{2.4} & 0.2 & 5.4 \\ \mathcal{L}'_V & 3.6 & \textbf{6.5} & 0.3 & 10.4 \\ \mathcal{L} \cap \mathcal{S} & 0.1 & \textbf{0.2} & 0.5 & 0.8 \\ \mathcal{S} & 11.0 & 4.5 & \textbf{79.0} & 94.4 \end{array} $	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$

94% of speech is hypothesized as speech

95% of laughter (excluding "laughed speech") is hypothesized as laughter

D . . . II

<回> < 回> < 回> < 回>

S 5.3 71.4 93.6

-

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○●○○○○	Summary

						Recall:	
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ		r coun	
\mathcal{N}	685.4	22.9	7.8	716.2			
\mathcal{L}'_U	2.8	2.4	0.2	5.4		\mathcal{L}	${\mathcal S}$
$\mathcal{L}_V^{}$	3.6	6.5	0.3	10.4	\mathcal{L}'	94.7	5.3
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8	$\mathcal{L}\cap\mathcal{S}$	28.6	71.4
${\mathcal S}$	11.0	4.5	79.0	94.4	${\mathcal S}$	5.4	93.6
Σ	702.9	36.6	87.8	827.2			

- 94% of speech is hypothesized as speech
- 95% of laughter (excluding "laughed speech") is hypothesized as laughter

Introduction	Data 00	Model 000000	Experiments 0000000	Analysis ○○○○●○○○	Summary

685.4

Σ	Precision:
716.2	

\mathcal{L}'_{II}	2.8	2.4	0.2	5.4	
$\mathcal{L}'_U \\ \mathcal{L}'_V$	3.6	6.5	0.3	10.4	\mathcal{L}'
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8	$\mathcal{L}\cap\mathcal{S}$
${\mathcal S}$	11.0	4.5	79.0	94.4	${\mathcal S}$
Σ	702.9	36.6	87.8	827.2	

S

7.8

	\mathcal{L}	${\mathcal S}$
	65.4	0.6
5	1.5	0.6
	33.1	98.8

-

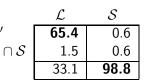
L

22.9

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○●○○○	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ	
\mathcal{N}	685.4	22.9	7.8	716.2	
\mathcal{L}'_U	2.8	2.4	0.2	5.4	
$\mathcal{L}'_U \ \mathcal{L}'_V$	3.6	6.5	0.3	10.4	\mathcal{L}'
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8	\mathcal{L}
S	11.0	4.5	79.0	94.4	\mathcal{S}
Σ	702.9	36.6	87.8	827.2	

Precision:



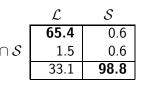
99% of hypothesized speech is speech

65% of hypothesized laughter is laughter

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○●○○○	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ	
\mathcal{N}	685.4	22.9	7.8	716.2	
\mathcal{L}'_U	2.8	2.4	0.2	5.4	
$\mathcal{L}_{V}^{}$	3.6	6.5	0.3	10.4	\mathcal{L}'
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8	\mathcal{L} (
S	11.0	4.5	79.0	94.4	\mathcal{S}
Σ	702.9	36.6	87.8	827.2	

Precision:



- 99% of hypothesized speech is speech
- 65% of hypothesized laughter is laughter

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○●○○	Summary

Confusions Between ${\cal L}$ and ${\cal N}$

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
\mathcal{L}'_V	3.6	6.5	0.3	10.4
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

\mathcal{N}	\mathcal{L}
685.4	22.9
2.8	2.4
3.7	6.7

.≣...>

- 4 ∃ ⊳

 $\bullet~$ looking at ${\cal L}$ and ${\cal N}~$ only

 $\mathcal{N} \\ \mathcal{L}'_U \\ \mathcal{L}_V$

Introduction	Data oo	Model 000000	Experiments	Analysis ○○○○○○●○	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
\mathcal{L}_{V}^{\prime}	3.6	6.5	0.3	10.4
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8
\mathcal{S}	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

Recall:

\mathcal{N}	\mathcal{L}
96.8	3.2
53.9	46.2
35.6	64.4

- 4 回 2 - 4 三 2 - 4 三 2

97% of silence is hypothesized as silence

64% of voiced laughter (including "laughed speech") is classified as laughter

 \mathcal{N} \mathcal{L}'_U \mathcal{L}_V

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○○●○	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
\mathcal{L}_{V}^{\prime}	3.6	6.5	0.3	10.4
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

Recall:

\mathcal{N}	\mathcal{L}
96.8	3.2
53.9	46.2
35.6	64.4

- * 同 * * ミ * * ミ *

э

97% of silence is hypothesized as silence

64% of voiced laughter (including "laughed speech") is classified as laughter

 \mathcal{N} \mathcal{L}'_U \mathcal{L}_V

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○○●○	Summary

						Recall:
	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ		
\mathcal{N}	685.4	22.9	7.8	716.2		
\mathcal{L}'_U	2.8	2.4	0.2	5.4		\mathcal{N}
$\mathcal{L}_{V}^{}$	3.6	6.5	0.3	10.4	\mathcal{N}	96.8
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8	\mathcal{L}'_U	53.9
${\mathcal S}$	11.0	4.5	79.0	94.4	\mathcal{L}_V	35.6
Σ	702.9	36.6	87.8	827.2		

\mathcal{N}	\mathcal{L}
96.8	3.2
53.9	46.2
35.6	64.4

向下 イヨト イヨト

- 97% of silence is hypothesized as silence
- 64% of voiced laughter (including "laughed speech") is classified as laughter

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○○●	Summary

Pre	CIC	non	٠
110	CIC	non	

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
\mathcal{L}'_V	3.6	6.5	0.3	10.4
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8
S	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

\mathcal{N}	\mathcal{L}
99.1	71.6
0.4	7.5
0.5	20.9

99% of hypothesized silence is silence

- 28% of hypothesized laughter is laughter
- 72% of hypothesized laughter is silence

 \mathcal{N} \mathcal{L}'_U \mathcal{L}_V

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○○○●	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ	
\mathcal{N}	685.4	22.9	7.8	716.2	
\mathcal{L}'_U	2.8	2.4	0.2	5.4	
\mathcal{L}'_V	3.6	6.5	0.3	10.4	\mathcal{N}
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8	\mathcal{L}'_U
${\mathcal S}$	11.0	4.5	79.0	94.4	\mathcal{L}_V^0
Σ	702.9	36.6	87.8	827.2	

Precision:

\mathcal{N}	\mathcal{L}
99.1	71.6
0.4	7.5
0.5	20.9

3 ×

99% of hypothesized silence is silence

28% of hypothesized laughter is laughter

3 72% of hypothesized laughter is silence

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○○○●	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ	
\mathcal{N}	685.4	22.9	7.8	716.2	
$\mathcal{L}_U' \ \mathcal{L}_V'$	2.8	2.4	0.2	5.4	
$\mathcal{L}_{V}^{'}$	3.6	6.5	0.3	10.4	\mathcal{N}
$\mathcal{L} \cap \mathcal{S}$	0.1	0.2	0.5	0.8	\mathcal{L}'_U
${\mathcal S}$	11.0	4.5	79.0	94.4	\mathcal{L}_V^0
Σ	702.9	36.6	87.8	827.2	

Precision:

\mathcal{N}	\mathcal{L}
99.1	71.6
0.4	7.5
0.5	20.9

- 99% of hypothesized silence is silence
- 28% of hypothesized laughter is laughter
- 3 72% of hypothesized laughter is silence

Introduction	Data 00	Model 000000	Experiments	Analysis ○○○○○○○●	Summary

	\mathcal{N}	\mathcal{L}	${\mathcal S}$	Σ
\mathcal{N}	685.4	22.9	7.8	716.2
\mathcal{L}'_U	2.8	2.4	0.2	5.4
$\mathcal{L}_{V}^{'}$	3.6	6.5	0.3	10.4
$\mathcal{L}\cap\mathcal{S}$	0.1	0.2	0.5	0.8
${\mathcal S}$	11.0	4.5	79.0	94.4
Σ	702.9	36.6	87.8	827.2

Precision:

\mathcal{N}	\mathcal{L}
99.1	71.6
0.4	7.5
0.5	20.9

- 99% of hypothesized silence is silence
- 28% of hypothesized laughter is laughter
- 3 72% of hypothesized laughter is silence

 \mathcal{N} \mathcal{L}'_U \mathcal{L}_V

Introduction	Data 00	Model 000000	Experiments	Analysis 00000000	Summary ●○○
Conclusion	IS				

Is baseline system for multiparticipant 3-way VAD

- no pre-segmentation assumed
- all laughter considered
- I aughter vs silence harder than aughter vs speech
- speech/laughter contrastive constraints helpful
 - maximum allowed degree of overlap
 - minimum state duration
- current performance is a function of
 - proportion of laughter present
 - participant novelty

(4回) (日) (日)

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	○○	000000	0000000	000000000	●○○
Conclusic	ons				

baseline system for multiparticipant 3-way VAD

- no pre-segmentation assumed
- all laughter considered

2 { laughter vs silence } harder than { laughter vs speech }
3 speech/laughter contrastive constraints helpful

maximum allowed degree of overlap
minimum state duration

3 current performance is a function of

proportion of laughter present

participant novelty

(4月) イヨト イヨト

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	oo	000000	0000000	000000000	●○○
Conclusic	ons				

baseline system for multiparticipant 3-way VAD no pre-segmentation assumed all laughter considered 2 { laughter vs silence } harder than { laughter vs speech } speech/laughter contrastive constraints helpful maximum allowed degree of overlap minimum state duration Current performance is a function of proportion of laughter present participant novelty

向下 イヨト イヨト

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	○○	000000	0000000	000000000	●○○
Conclusio	ons				

baseline system for multiparticipant 3-way VAD

- no pre-segmentation assumed
- all laughter considered
- { laughter vs silence } harder than { laughter vs speech }
- speech/laughter contrastive constraints helpful
 - maximum allowed degree of overlap
 - minimum state duration
- current performance is a function of
 - proportion of laughter present
 - participant novelty

向下 イヨト イヨト

Introduction	Data	Model	Experiments	Analysis	Summary
0000000	oo	000000	0000000	000000000	●○○
Conclusic	ons				

- Description of the system for multiparticipant 3-way VAD
 - no pre-segmentation assumed
 - all laughter considered
- { laughter vs silence } harder than { laughter vs speech }
- speech/laughter contrastive constraints helpful
 - maximum allowed degree of overlap
 - minimum state duration
- current performance is a function of
 - proportion of laughter present
 - participant novelty

Introduction	Data 00	Model 000000	Experiments	Analysis 00000000	Summary ○●○
Possible I	⁻ uture W	ork			

() model voiced and unvoiced laughter $(\mathcal{L}_V \text{ and } \mathcal{L}_U)$ separately

- different acoustics
- different overlap contexts (Laskowski & Burger, 2007c)
- different semantics

② characterize laughter by instrumentality to high level tasks

- which laughter signals different emotional valence
- which laughter signals involvement hotspots
- Q: Does instrumentality correspond to how clear and unambiguous laughter is? cf. (Truong & van Leeuwen, 2007a)
- multi-pass, multi-resolution laughter detection
 - pass 1: large frame size (0.1s), small context (0.1s)
 - pass 2: small frame size (0.01s), large context (1.0s) (Knox & Mirghafori, 2007)

(日) (四) (王) (王)

Introduction	Data oo	Model 000000	Experiments	Analysis	Summary ○●○
Possible I	- uture W	ork			

() model voiced and unvoiced laughter $(\mathcal{L}_V \text{ and } \mathcal{L}_U)$ separately

- different acoustics
- different overlap contexts (Laskowski & Burger, 2007c)
- different semantics

② characterize laughter by instrumentality to high level tasks

- which laughter signals different emotional valence
- which laughter signals involvement hotspots
- Q: Does instrumentality correspond to how clear and unambiguous laughter is? cf. (Truong & van Leeuwen, 2007a)
- Implication multi-resolution laughter detection
 - pass 1: large frame size (0.1s), small context (0.1s)
 - pass 2: small frame size (0.01s), large context (1.0s) (Knox & Mirghafori, 2007)

(日) (四) (王) (王)

Introduction	Data oo	Model 000000	Experiments	Analysis	Summary ○●○
Possible I	Future W	ork			

() model voiced and unvoiced laughter $(\mathcal{L}_V \text{ and } \mathcal{L}_U)$ separately

- different acoustics
- different overlap contexts (Laskowski & Burger, 2007c)
- different semantics

2 characterize laughter by instrumentality to high level tasks

- which laughter signals different emotional valence
- which laughter signals involvement hotspots
- Q: Does instrumentality correspond to how clear and unambiguous laughter is? cf. (Truong & van Leeuwen, 2007a)
- Image: multi-resolution laughter detection
 - pass 1: large frame size (0.1s), small context (0.1s)
 - pass 2: small frame size (0.01s), large context (1.0s) (Knox & Mirghafori, 2007)

イロト イヨト イヨト イヨト

Introduction	Data oo	Model 000000	Experiments	Analysis	Summary ○●○
Possible I	Future W	ork			

() model voiced and unvoiced laughter (\mathcal{L}_V and \mathcal{L}_U) separately

- different acoustics
- different overlap contexts (Laskowski & Burger, 2007c)
- different semantics

2 characterize laughter by instrumentality to high level tasks

- which laughter signals different emotional valence
- which laughter signals involvement hotspots
- Q: Does instrumentality correspond to how clear and unambiguous laughter is? cf. (Truong & van Leeuwen, 2007a)
- Image: multi-resolution laughter detection
 - pass 1: large frame size (0.1s), small context (0.1s)
 - pass 2: small frame size (0.01s), large context (1.0s) (Knox & Mirghafori, 2007)

- 4 回 ト 4 ヨ ト 4 ヨ ト

Introduction	Data oo	Model	Experiments	Analysis 000000000	Summary ○○●		
Thanks for attending							

Also, thanks to

- \bullet Susi Burger, help with ${\cal L}$ segmentation & classification
- Liz Shriberg, access to ICSI MRDA Corpus
- Khiet Truong and Mary Knox, discussion of own work