Engaging Content
Engaging People

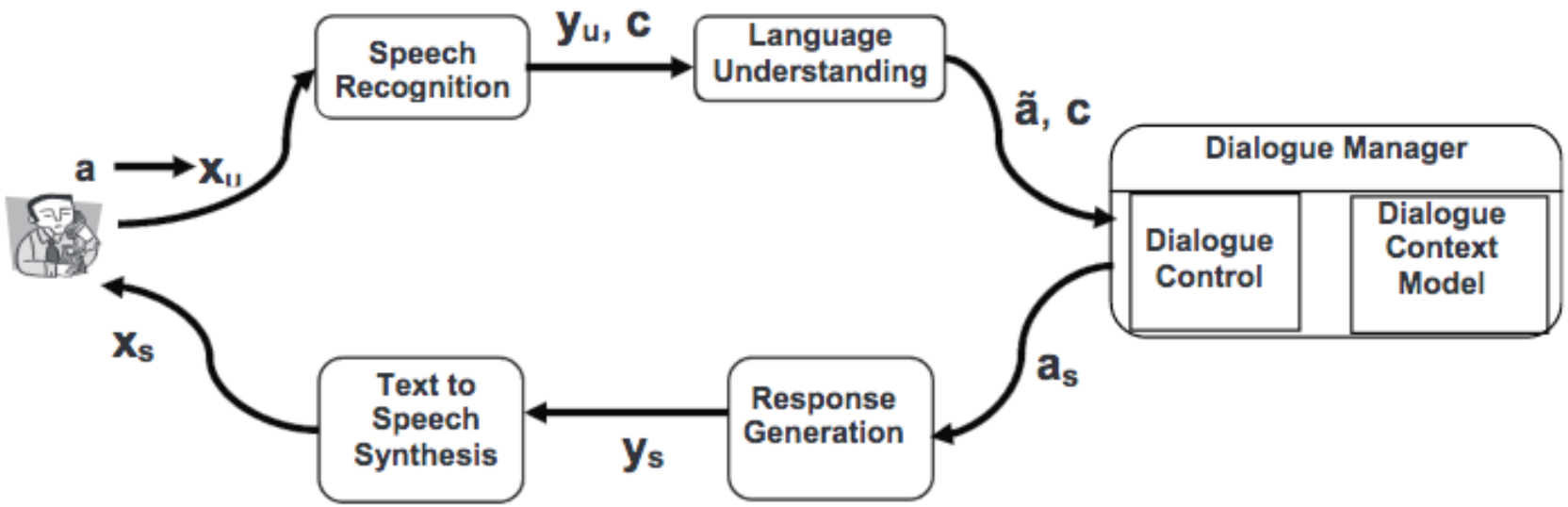
Talking to Machines: Conversation

Emer Gilmartin,
ADAPT Centre
Trinity College Dublin

Outline

- Current Situation
- Future Conversations
  - Instrumental vs Interactive talk
  - Casual Conversation Structure
  - ADELE Corpus - Greeting and Leavetaking
  - Multiparty Chat and Chunk modelling
- Other considerations
  - ASR
  - TTS
  - Multimodality

Spoken Dialog System



$\mathbf{x}_u$ – user acoustic signal	$\mathbf{a}_s$ – system dialogue act
$\mathbf{y}_u$ – speech recognition hypothesis (words)	$\mathbf{y}_s$ – system word string
a – user dialogue act (intended)	$\mathbf{x}_s$ – system acoustic signal
ã – user dialogue act (interpreted)	

What?

- Spoken dialogue systems attempt to create a spoken interaction with a user
- Dialogue systems
  - Intelligent Virtual Agents (IVA's), Embodied Conversational Agents (ECA's), Chatbots
- Dream (Turing, 1950) vs Practical Progress (Allen, 2000)
  - AI – early chat – pattern matching – ELIZA
  - Practical Dialogues – task to be performed - Practical Dialogue Hypothesis (Allen, 2000)

What's out there?

- Command and Control – voice commands
- Interactive Voice Response – IVR
- Information Retrieval – voice search
- Siri, Alexa, Google Home
- Chatbots
- Embodied Conversational Agents (ECA)
- Intelligent Virtual Agents

The Problem: Building social dialogue systems entails understanding of casual social dialogue but…

- Much linguistic theory is based on language similar to writing but highly unlike talk
  - regards spoken interaction as debased, chaotic
- SDS technology based on
  - Practical Dialogue Hypothesis (Allen, 2000)
  - Constraint introduced to make dialogue modelling tractable
- Much corpus study of spoken interaction based on Task-based Dialogue
  - Information gap activities – MapTask (HCRC), DiaPix (Lucid)
  - Meetings – AMI, ICSI
  - These are not corpora of casual or social talk

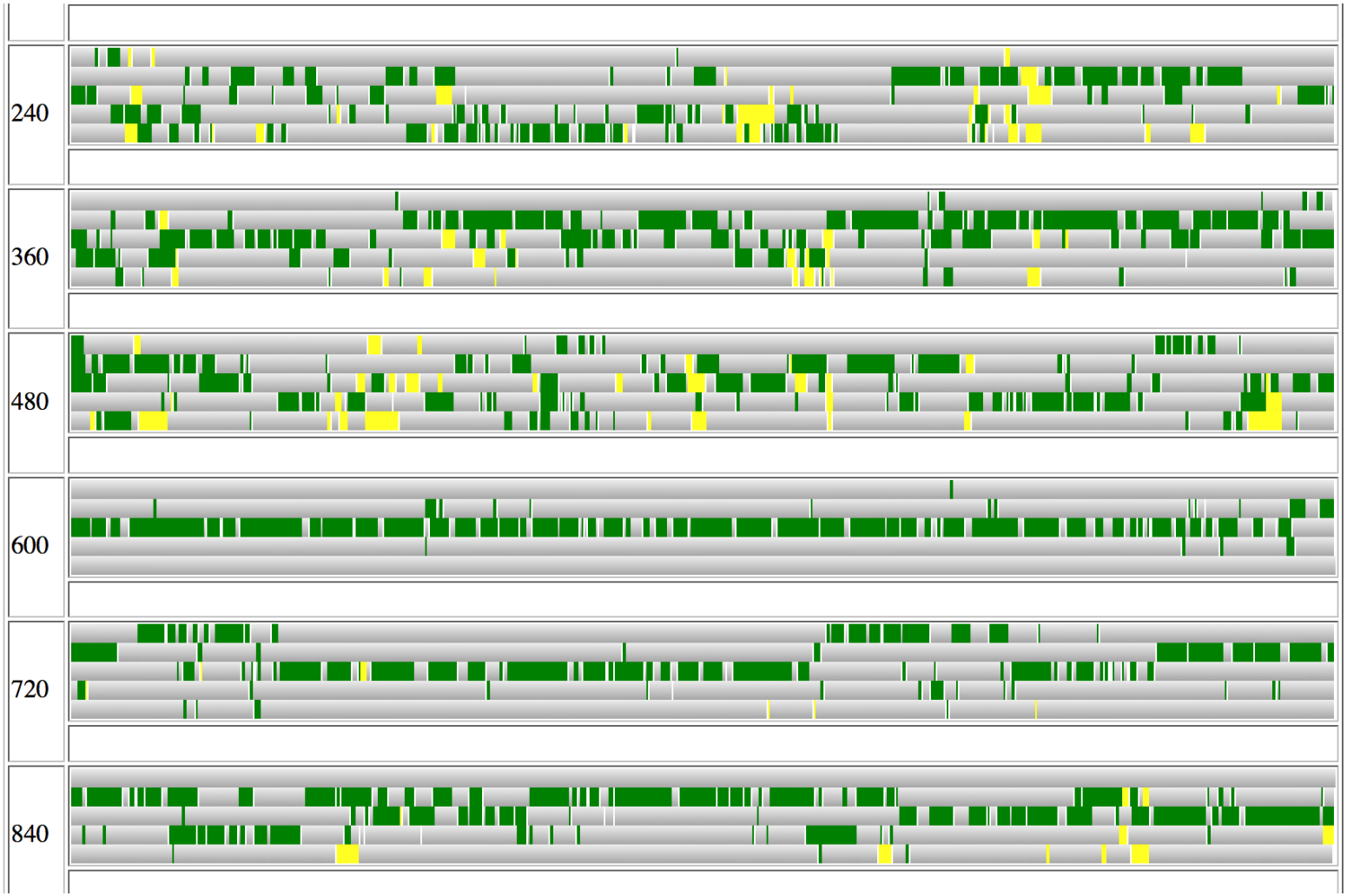
Transactional v Interactional Conversation

- Ordering a pizza (transactional)
  - performing a well-defined task
  - content (‘What?’) vital for success
- Chat with neighbour (interactional)
  - building/maintaining social bonds
  - social (‘How?’) very important
- Longer form (c 1 hr) casual conversation
  - ‘continuing state of incipient talk’
- Growing interest in interactional conversations

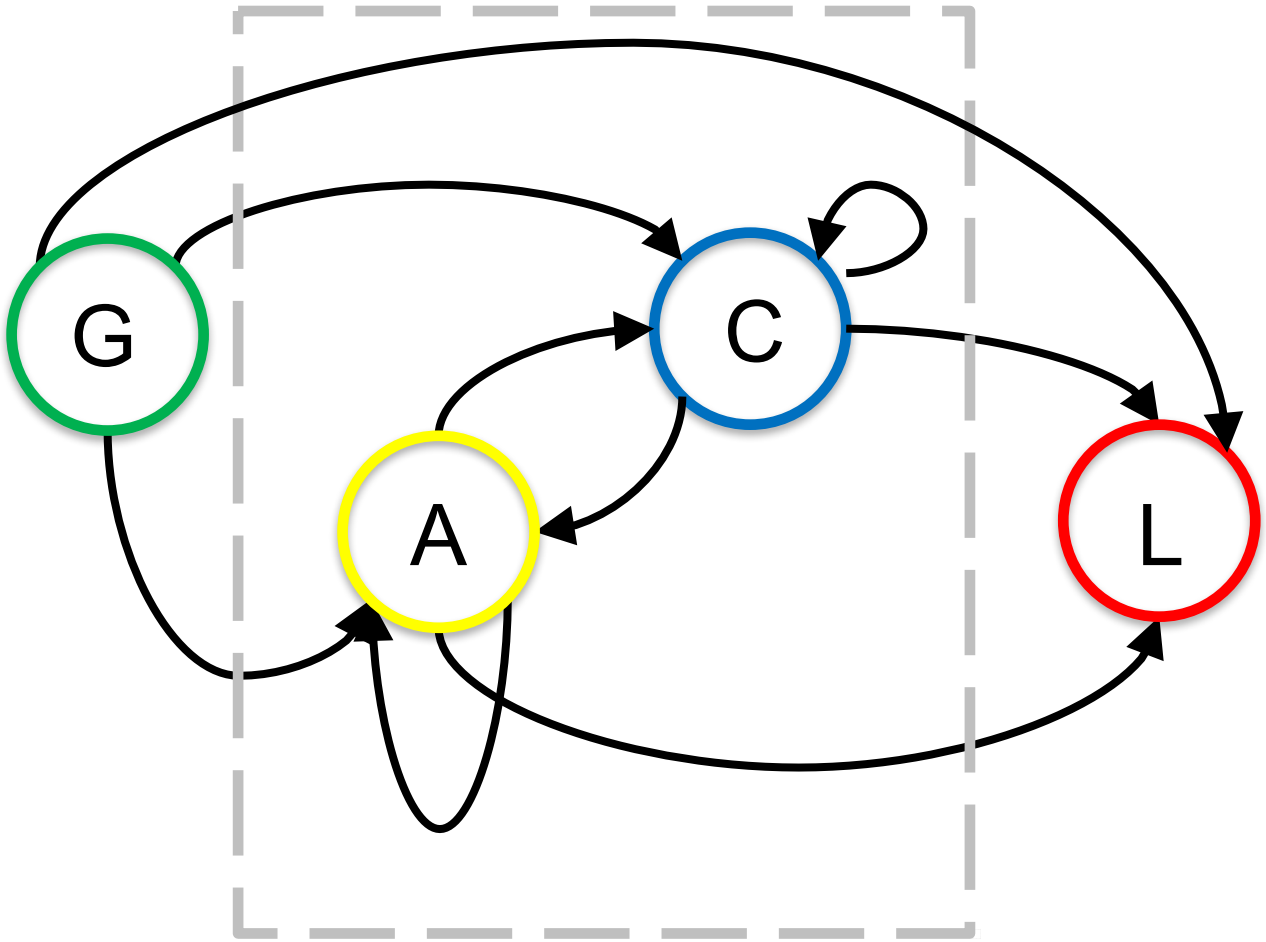
Social / Casual Talk

- Spoken interaction as social activity
  - Malinowski, Dunbar, Jakobsen, Brown and Yule
- Structure and Content
  - Smalltalk at the margins (Laver)
  - Chat and chunks (Slade & Eggins)
    - chat – highly interactive, many speakers contributing
    - chunks – gossip, narrative, dominated by one speaker
  - Phases – greetings, approach, centre, leavetaking (Ventola)
  - Multiparty (Slade)
- Problems:
  - much of this is theory, analysis by example
  - based on orthographical transcriptions
  - corpus based studies on transactional dyadic interaction, phonecalls…

12 minutes from a 5-party casual conversation showing chat (240s-480s and chunk 480 – end) phases
Green-speech, yellow-laughter, grey-silence



Anatomy of casual conversation (Ventola model)



Genre differences in spoken interaction?

- Spoken interaction is situated
  - ‘speech-exchange systems’ (SSJ),
  - communicative activities (Allwood)

- Some low level mechanisms may follow universal patterns
- It is also possible that even basic interaction mechanisms such as turn-taking vary with the type and parameters of different interactions

- What might vary?
  - Utterance/turn characteristics
  - Distribution of pauses/gaps/overlaps
  - ‘Disfluencies’, VSU’s, laughter…

- Explore different genres and use knowledge to inform design of interfaces

Engaging Content
Engaging People

Annotation of Greeting and Leave-taking in Social Text Dialogues Using ISO 24617-2

Emer Gilmartin, Brendan Spillane, Maria O'Reilly, Christian Saam, Ketong Su, Killian Levacher, Loredana Cerrato, Benjamin R. Cowan, Leigh M. H. Clark, Arturo Calvo, Nick Campbell, Vincent Wade

ADELE Corpus

- Purpose
  - Training data for SDS
- Scenario
  - Dyadic text interaction
- Data Collection
  - 37 participants (26M/11F, age range 18-43)
  - native English speakers or IELTS 6.5
  - working/studying and living in Ireland
  - 193 completed dialogues were collected.
- Data
  - 40,297 words over 9231 turns or ‘utterances’ (~200, 50)
  - 7811 or 84.7% tagged with a single label
  - 1209 (13%) - two tags, 181 (2%) - three tags
  - 26 (0.3%) and 3 utterances had four and five tags respectively.

Annotation of social acts

- Many schemes include social acts
- In a survey of 14 schemes, Petukova found
  - 10 included greeting functions, 4 included introductions, 6 had goodbyes, 5 included apology type functions, and 5 contained thanking
- The Social Obligations Management dimension of the ISO standard contains nine communicative functions
  - initialGreeting, initialSelfIntroduction, returnSelfIntroduction, apology, acceptApology, thanking, acceptThanking, initialGoodbye, and returnGoodbye.

Annotation

- Used ISO Standard (with additions)
- Lexical tags for topic – PropQuestion[hobby]
- Informs that were not first mentions tagged as comments
- Noticed problems with SOM – greetings, introductions, leavetaking
- Greeting sections were marked as beginning with the first utterance of the conversation, and ending with the last production of a formulaic greeting/introduction or greeting/introduction response.
- leave-taking sequences from the first attempt to close the conversation to the final utterance of the conversation.

Additional GIL Acts

Table 1: Acts introduced for the ADELE annotation and common surface forms

Act	Common Examples	Functional Area
ntmy	Nice to meet you	Greeting
	Good to talk to you	Greeting
repNtmy	Nice to meet you too	Greeting
	Good to talk to you too	Greeting
hay	How are you?	Greeting
	How's it going?	Greeting
repHay	Fine	Greeting
greet	Hello	Greeting
	Hi	Greeting
wntmy	It was lovely to meet you	leave-taking
	Nice talking to you	leave-taking
repWntmy	It was nice to meet you too	leave-taking
	Likewise	leave-taking

Distribution of GIL acts

Table 2: Greeting, Introduction, and Leavetaking (GIL) Acts in ADELE corpus

Description	Count	% Corpus
All acts included in GIL sequences (GILseq)	2336	21.5
GILA: Only GIL Acts: GILseq Acts - Interloper Acts	1820	16.7
GILB: Only GIL acts without LeaveTaking Introductions: GILA - Leavetaking Introductions	1626	15
Social Obligation Management Acts (SOM) other than GIL	198	2

Future: Contributing to revised ISO

Engaging Content
Engaging People

Exploring Multiparty Casual Talk for Social Human-Machine Dialogue

Genre differences in spoken interaction?

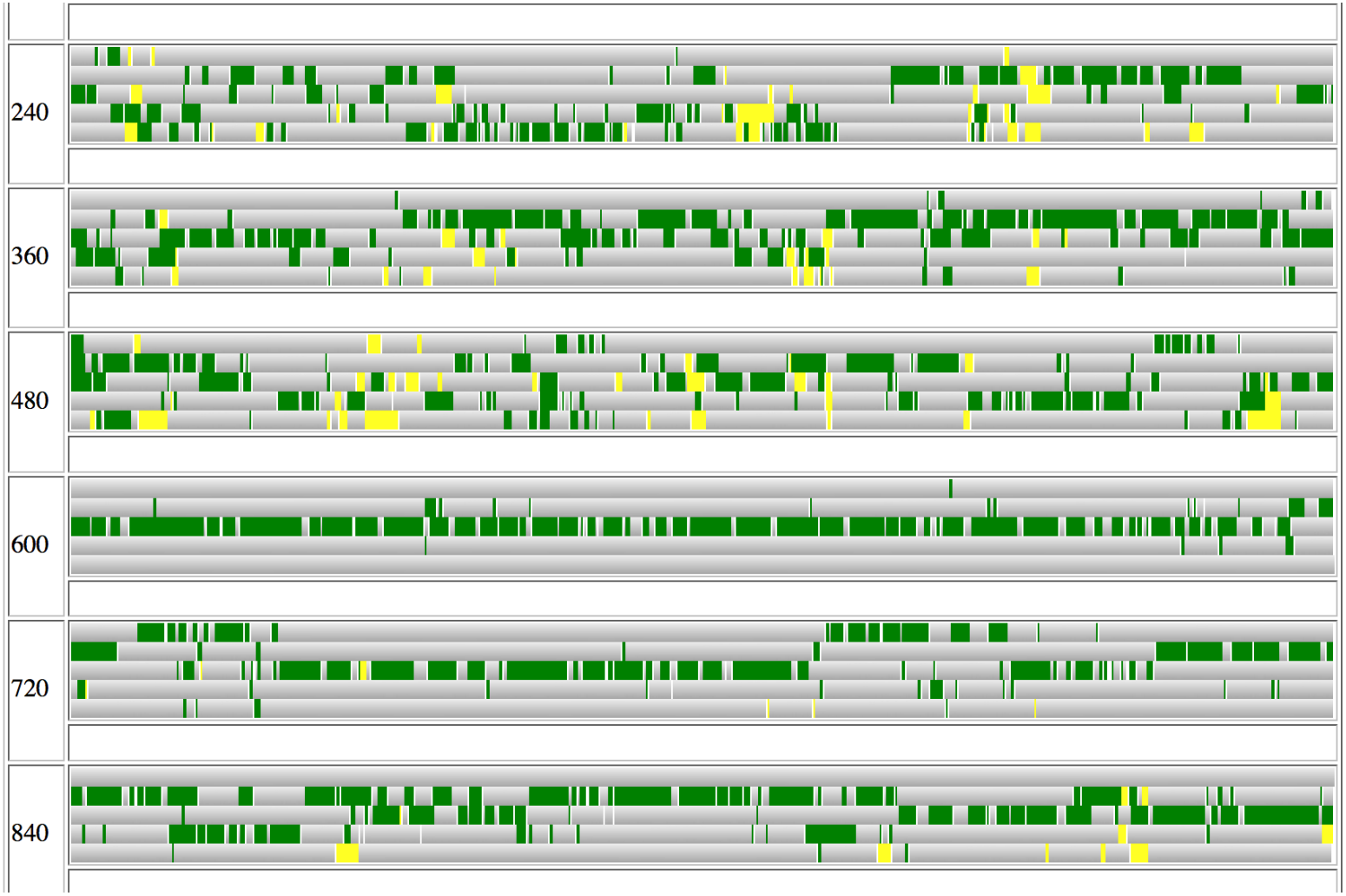
- Spoken interaction is situated
  - ‘speech-exchange systems’ (SSJ),
  - communicative activities (Allwood)

- Some low level mechanisms may follow universal patterns
- It is also possible that even basic interaction mechanisms such as turn-taking vary with the type and parameters of different interactions

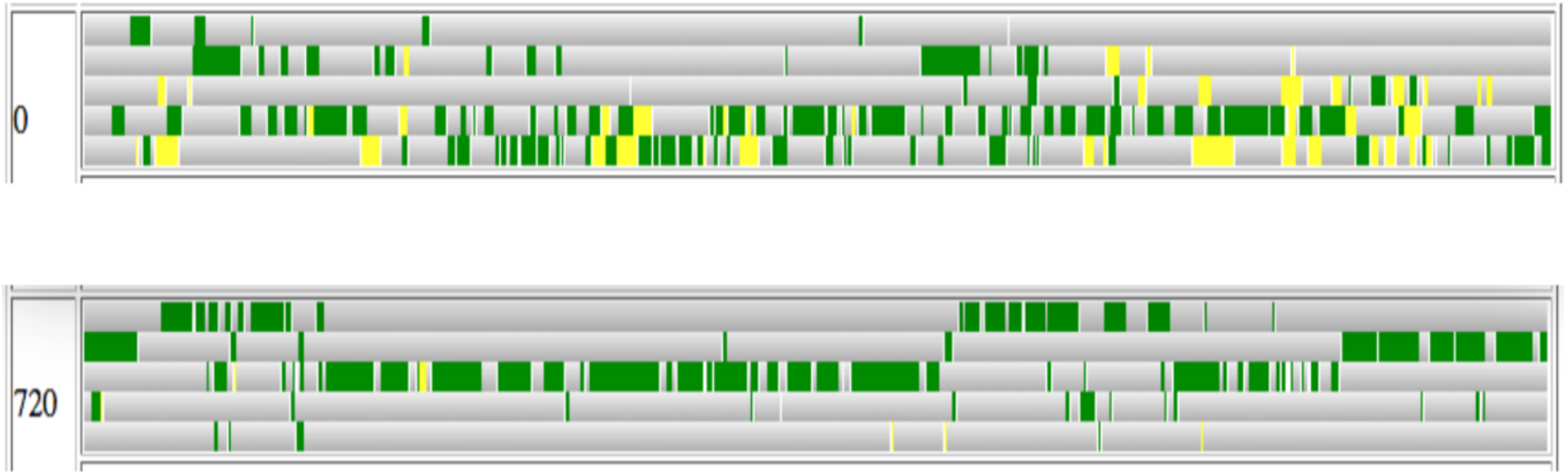
- What might vary?
  - Utterance/turn characteristics
  - Distribution of pauses/gaps/overlaps
  - ‘Disfluencies’, VSU’s, laughter…

- Explore different genres and use knowledge to inform design of interfaces

12 minutes from a 5-party casual conversation showing chat (240s-480s and chunk 480 – end) phases
Green-speech, yellow-laughter, grey-silence



Chat and Chunk



Question

Can chat and chunk phases be classified using acoustic/discourse features?

Data and annotation

Corpus	Participants	Gender	Duration (s)
D64	5	2F/3M	4164
DANS	3	1F/2M	4672
DANS	4	1F/3M	4378
DANS	3	2F/1M	3004
TableTalk	4	2F/2M	2072
TableTalk	5	3F/2M	4740

Table 1. Source corpora and details for the conversations used in dataset

Chat/Chunk Results

Significant differences in:

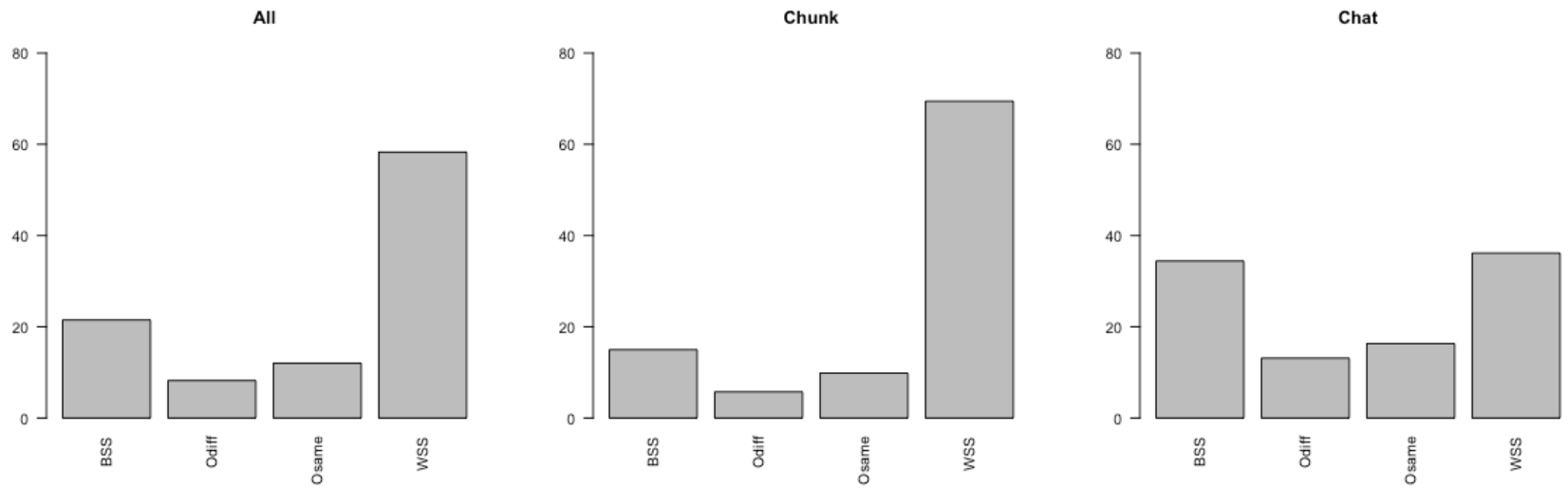
- Length – (chat more variable) gmean ~ 28s, chunk ~ 30s
- Distribution, more chat at beginning – c.8 minutes
- Laughter – over twice as much in chat – 9.7 vs 4%
- Gap lengths and distribution – WSS most common overall, more BSS in chat
- Overlap – more in chat, particularly more multiparty overlap
- Disfluency distribution, especially fp in chunks by role

Overlap and gap results

Speaker change: Between speaker silence (BSS) and between speaker overlap (Odiff)

Turn retention: Within speaker silence (WSS) and within speaker overlap (Osame)

Distributions differ between chunk and chat



Discussion

Important because;

- Need different timing modules for different phases
  - Many within speaker pauses in chunks are longer than between speaker pauses in chat so need different turntaking policies
- Suit different tasks – companion applications
  - System can recognise when to listen to a story (chunk)
- Aid comprehension – design educational dialogue in chunks

Current and Future Work

- Stochastic model
  - Preliminary results promising
- Goals
  - online classifier
  - incorporate in social dialogue system. CALL applications

Other considerations

- Voice

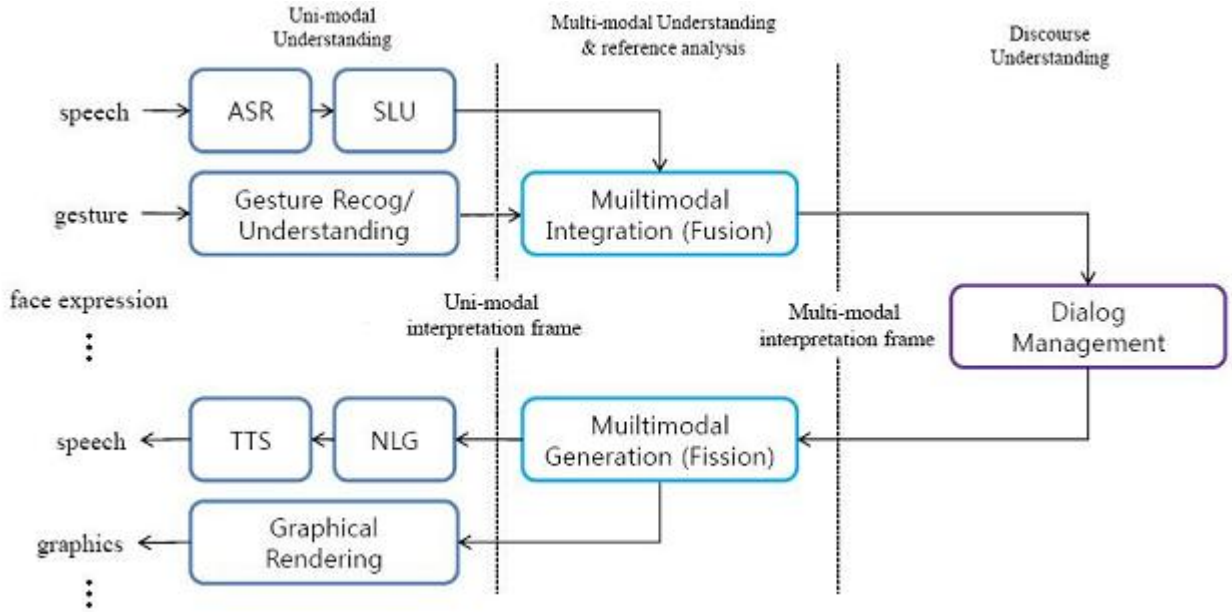
- Turn management / Endpointing
- Conversational ASR not there yet.

Multimodality

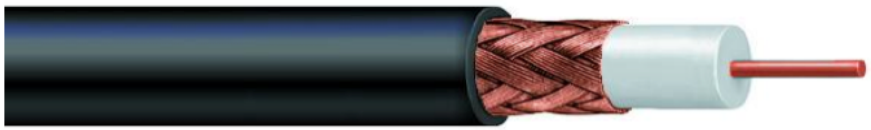
Expression and Recognition

Audio, visual, verbal, vocal, non-verbal, facial expression, gesture, posture…

Presence, affect, attitude...



Spoken interaction is more than just words!



To better understand and model the bundle of signals in conversation

Engaging Content
Engaging People

Thank You

Questions?