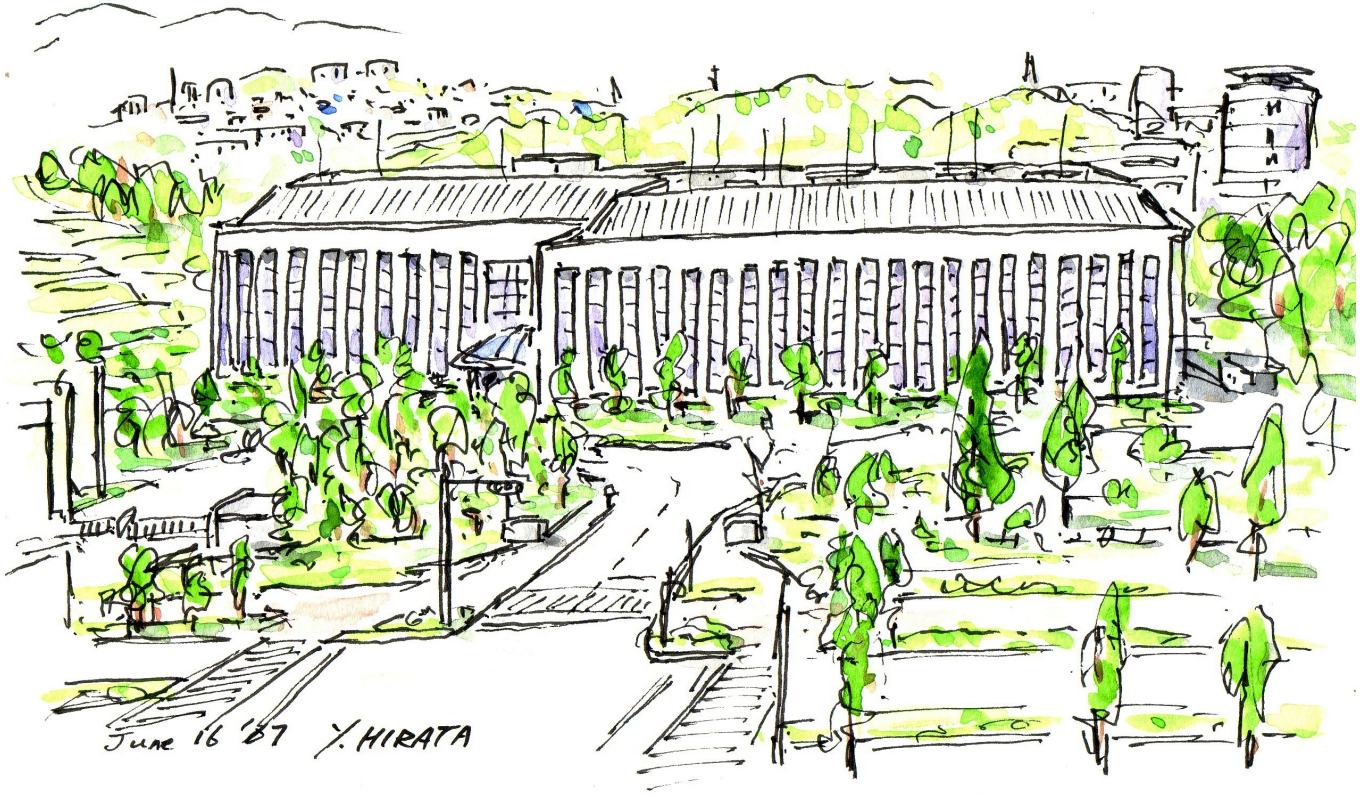
Tools & Resources for Visualising Conversational-Speech Interaction

Nick Campbell

NiCT/ATR-SLC
Keihanna Science City,
Kyoto, Japan.
nick@nict.go.jp

Greetings from the NiCT/ATR labs in Kyoto (from nick@nict.go.jp)



Introduction

With ever-growing increases in the amount of data available for speech technology research, it is now increasingly difficult for any one individual to become personally familiar with all of the data in any given corpus.
Yet without the insights provided by first-hand inspection of the types and variety of speech material being collected, it is difficult to ensure that appropriate models and features are being used in the processing of the speech data.

Introduction (cont.)

It is perhaps not necessary (and often physically very difficult) to listen to all of the speech in a given corpus but it is essential to be able to select in a non-random manner specific sections of the corpus for closer inspection and analysis.
If the data is transcribed, the transcriptions will provide the first key into the speech data but there are many aspects of a spoken message that are not well described by a plain text rendering of the linguistic content.

The JST/ATR "ESP" Corpus

- 5-years of recordings
- 50 paid volunteer speakers
- close-talking head-mounted mics worn throughout the day
- 1,500 hours of manually transcribed conversational speech
- = a lot of noise to analyse!!

The ATR SCOPE/Kaken Dialogue Corpora

- 3-years of dialogue recordings
- both camera & microphone recordings
- multi-person business and social interactions
- also annotated for discourse moves
- = a lot of data to associate!!

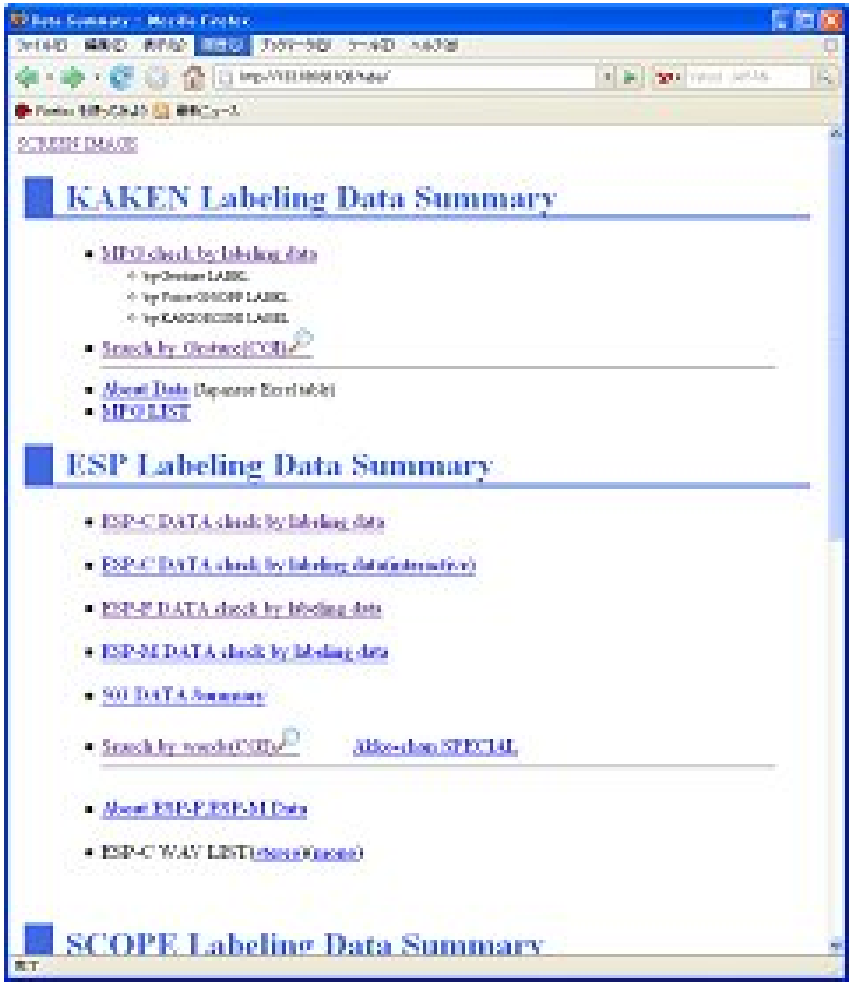
New ways to access the data

- an interactive web-based interface
- viewing dialogue structure - patterns of interaction
- accessing movement data & aligning it with speech segments
- finding synchrony in discourse participation
- measuring “rapport” in a conversation

Hierarchical structure to the data & transcriptions

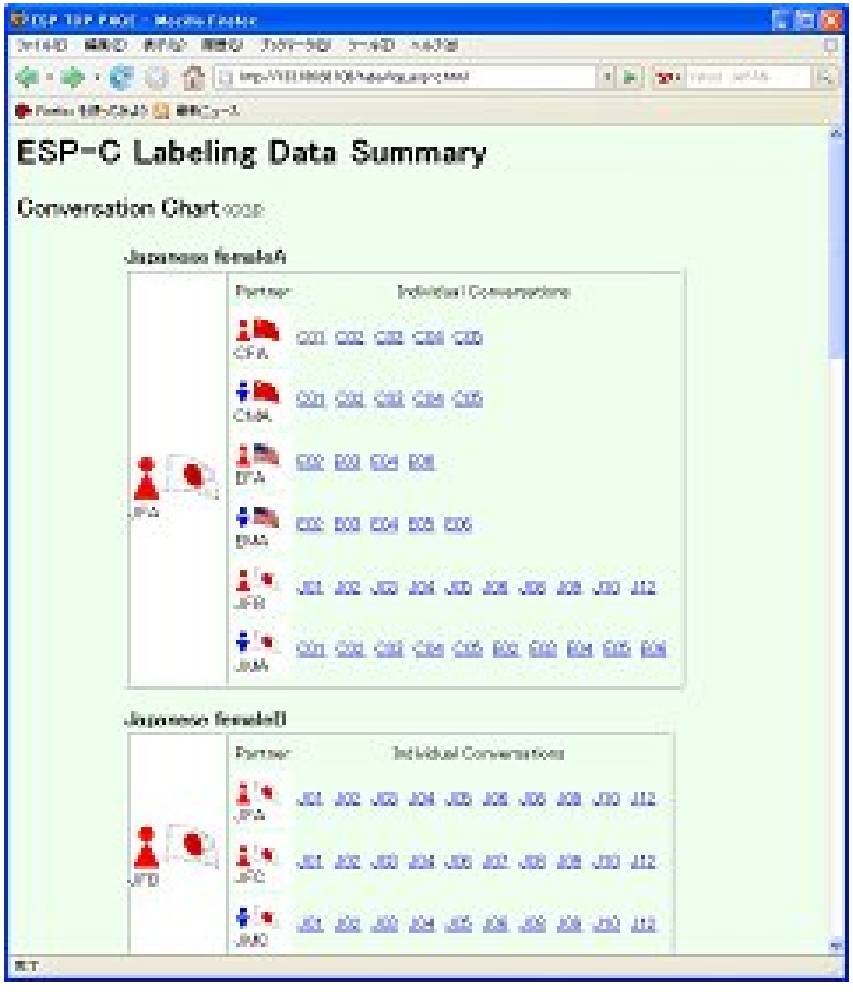
- all transcriptions are (manually) time-aligned
- multiple views of the same discourse in separate directories
- all files linked through cgi scripts using timing infomation
- accessed through ordered search
- new annotations linked as data grows and is shared

http://feast.atr.jp : top



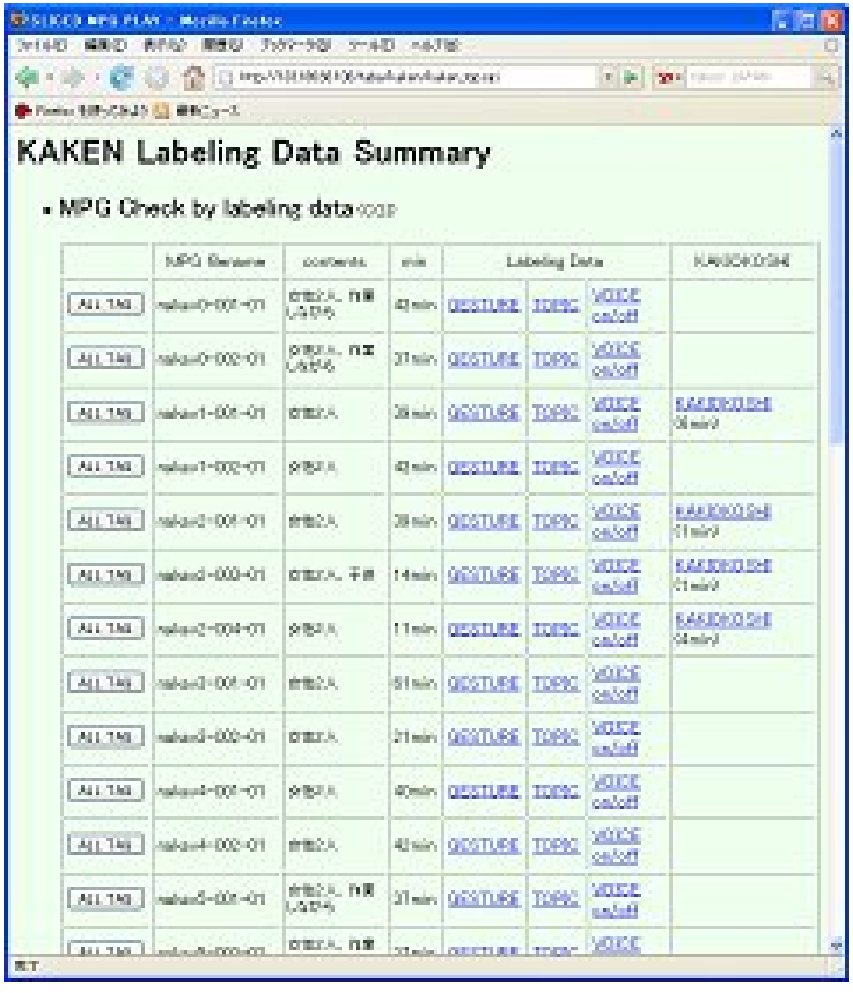
FEAST: Feature Extraction & Analysis for Speech Technology

The top-level web-pages for ESP_C



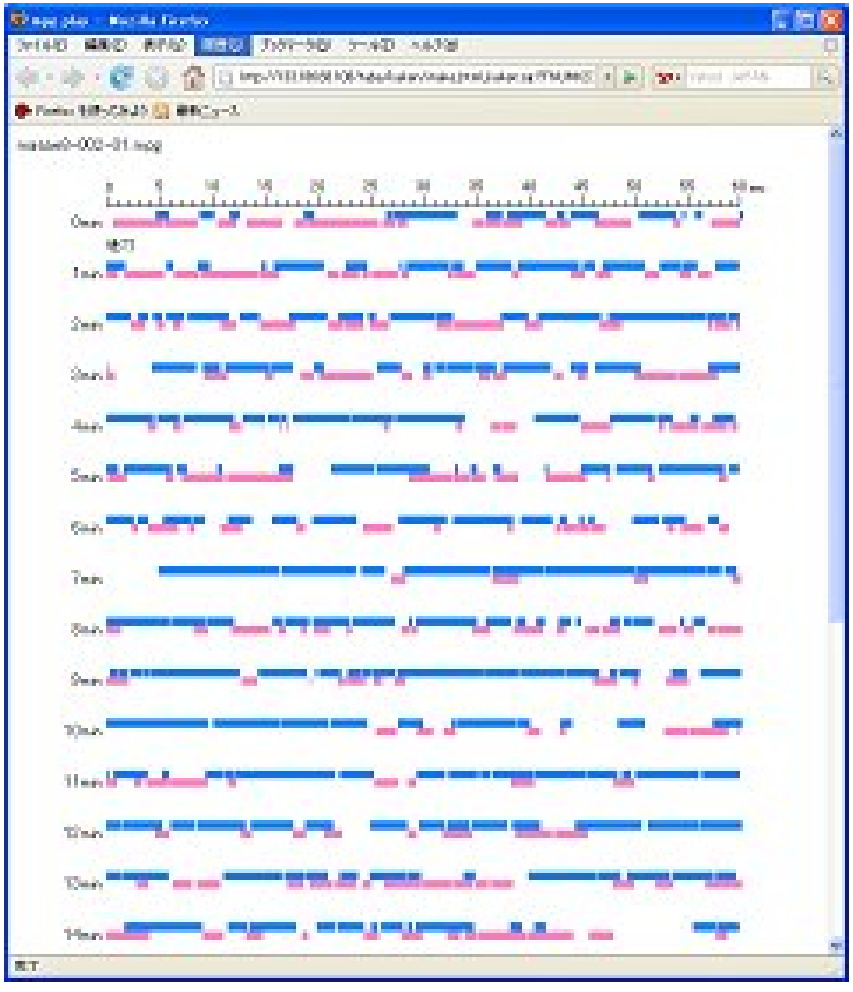
ESP-C: a 100-hour telephone conversation corpus

Multimodal Dialogue Data (with Kobe University)



Multimodal data - annotated video & audio

Visualising a conversation

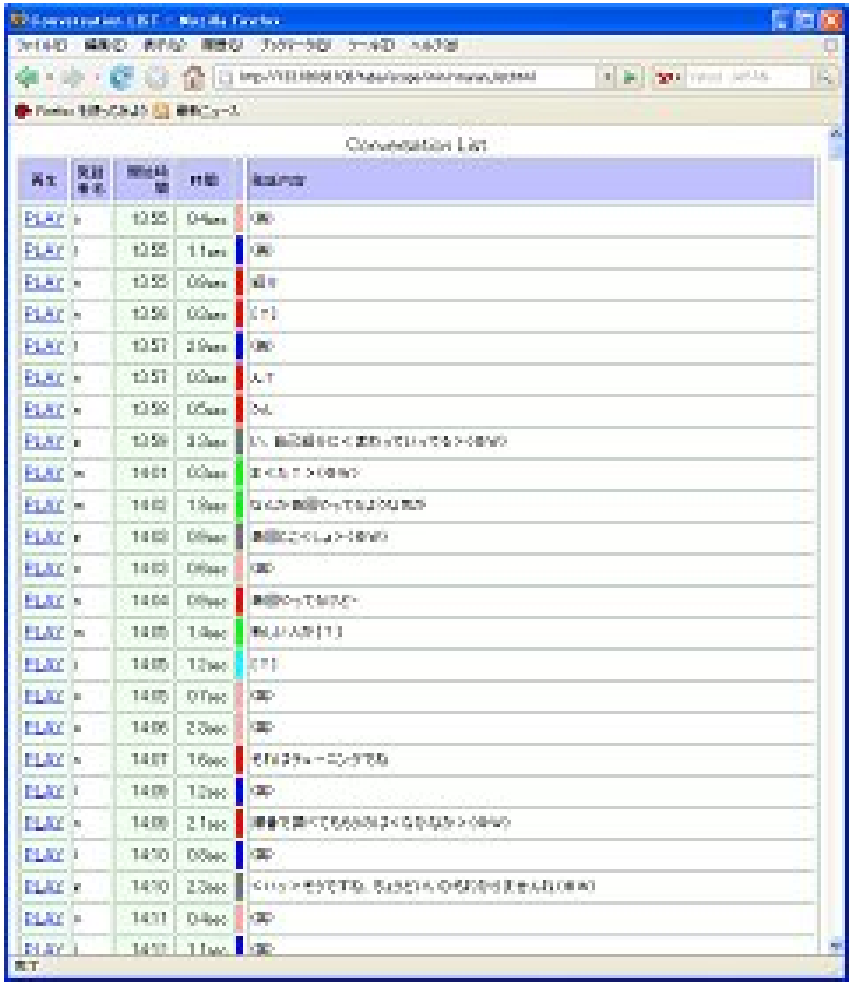


Speech activity for the two participants, showing overlaps & turns

View of a Multi-Party Dialogue

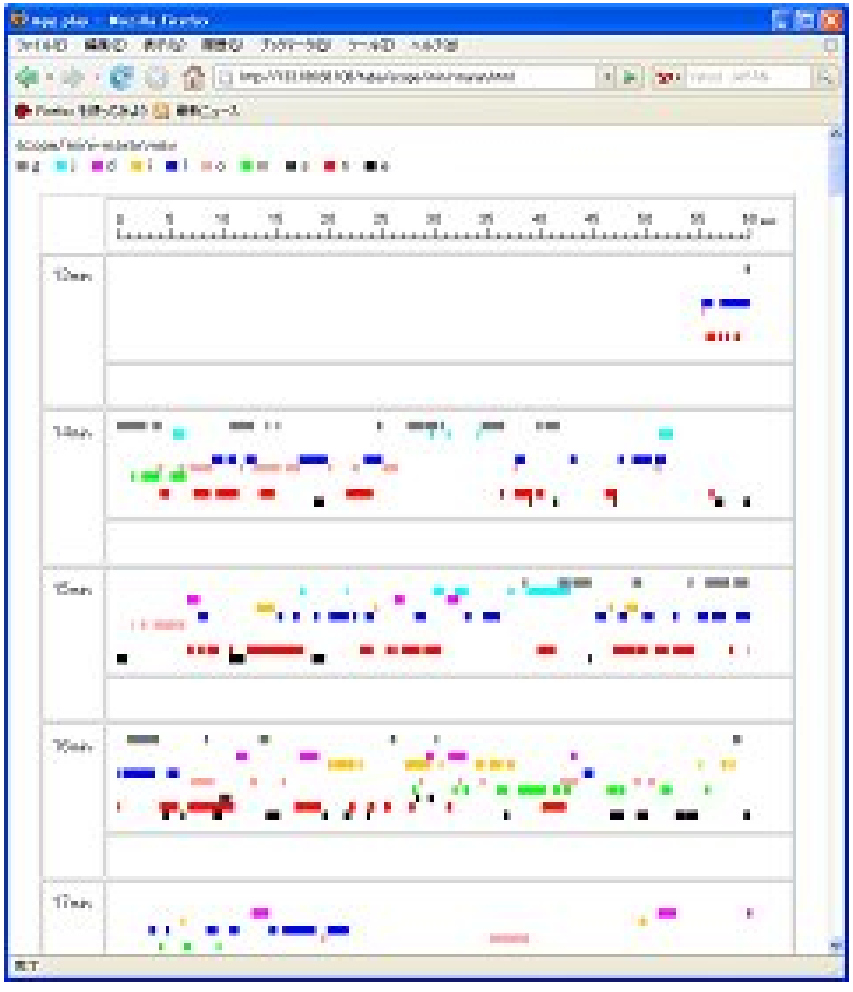
In a multi-party discourse, each speaker is colour-coded

Text of a multi-person dialogue



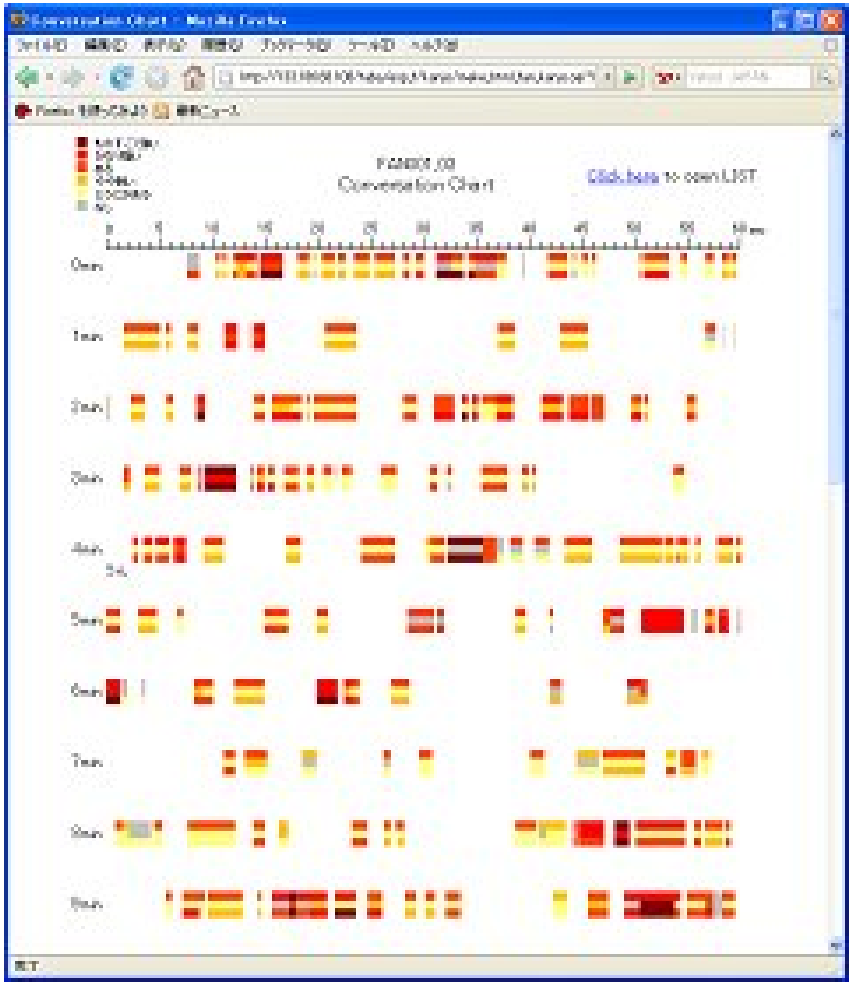
Transcription lines are clickable - each can be listened to

Display of a multi-person dialogue



Colour-coded bar-plots allow a better view of the interaction

Labelling the "intensity" of a discourse (rapport)



Three labellers indicate discourse involvement for 1 speaker

Viewing the video clip per utterance segment

By clicking on a bar, its video sequence can be viewed

Plotting head and body motion

Or movement data plotted using the same colour-coding

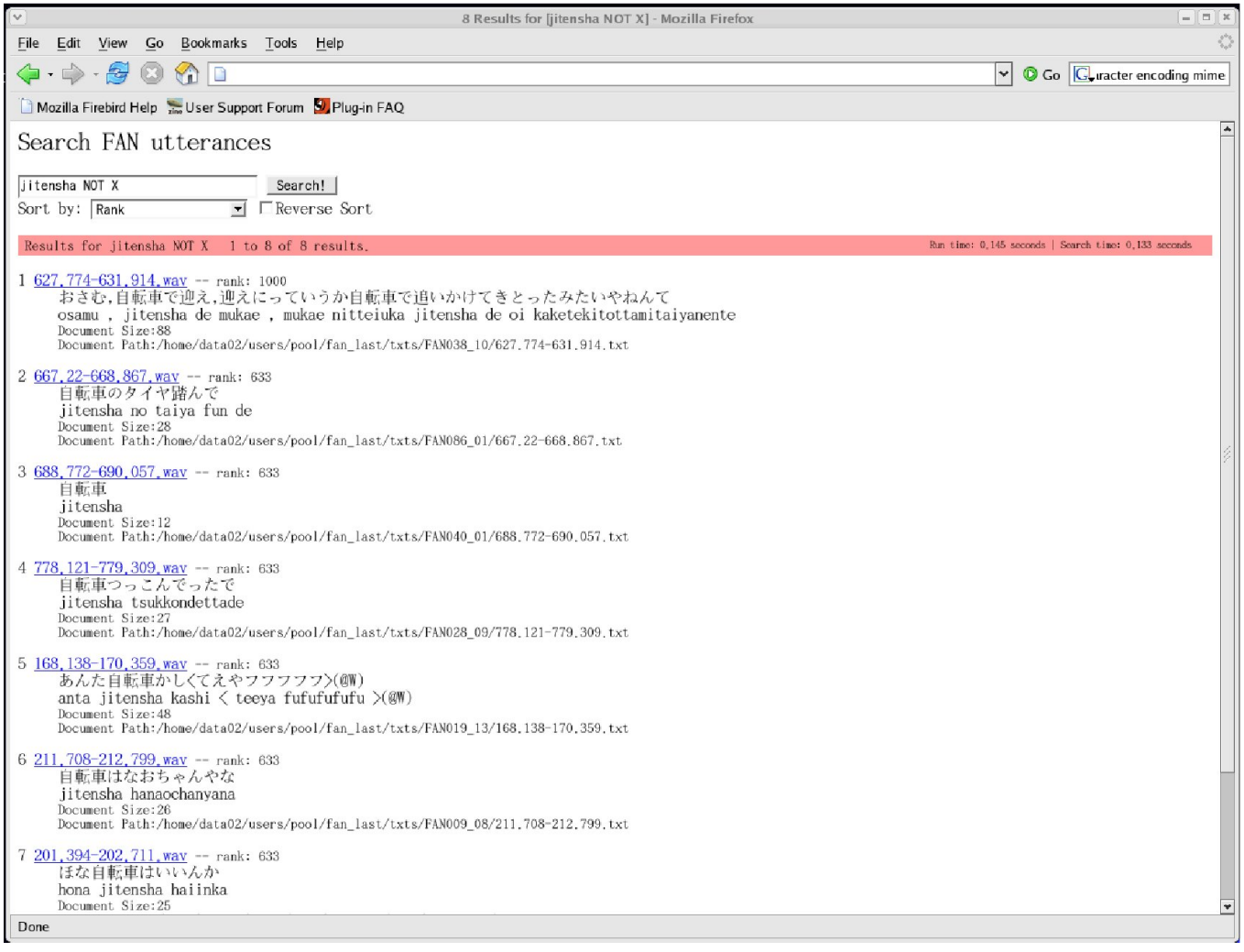
Gesture

individual examples of a given gesture can be listed for viewing

Searching by Gesture Label

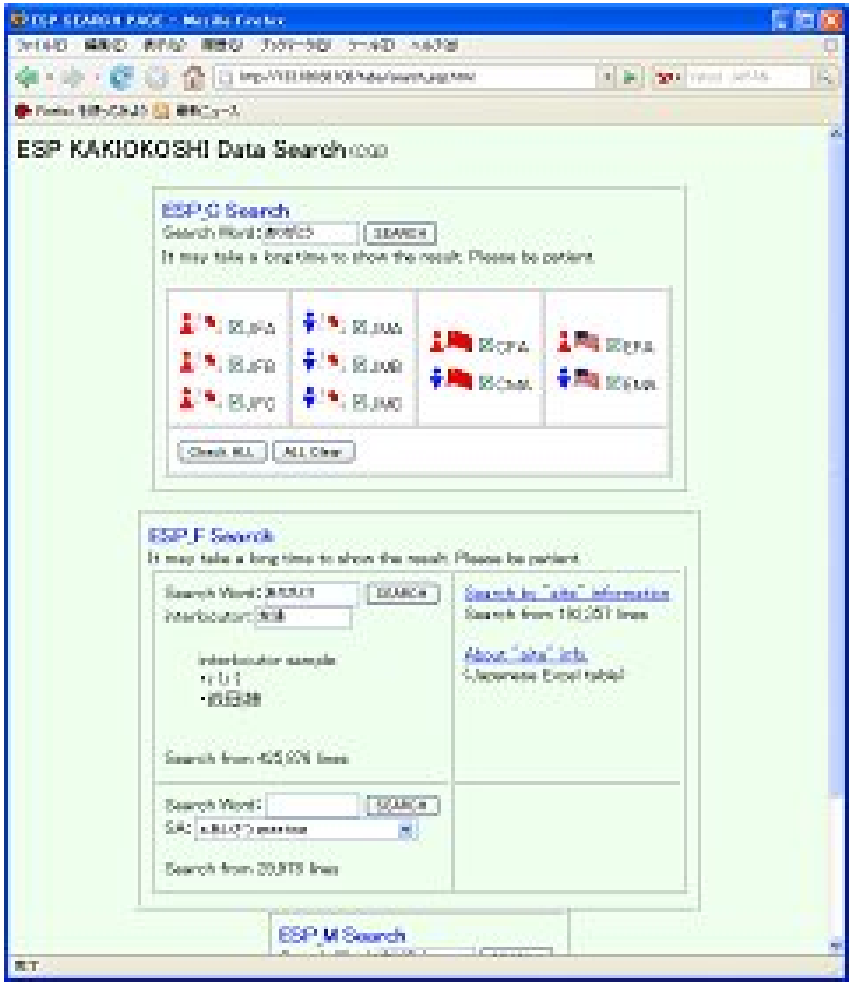
Gestures have been labelled and can be searched for by type

SWISH - a Google-type interface to large text data



A fast search-engine interface is included for text-based search

Feature-based (targeted) search



Search conditions can be specified in a discourse-specific way

Multinational English Conversational Data

Conversation Data in English

MPG data with subtitles '07/11/05 – '07/11/07

DAY 1

DAY 2

DAY 3

Conversation Chart & LIST

data	length		CHART	Emotion CHART		LIST	TOPIC LIST	MPG data
DAY 1	34:35	4 persons with FACE trace	CHART	labeller A	labeller B	LIST	TOPIC	DOWNLOAD
DAY 2	01:22:15	5 persons	CHART	labeller A	labeller B	LIST	TOPIC	DOWNLOAD
DAY 3	01:22:45	4 persons	CHART	labeller A	labeller B	LIST	TOPIC	DOWNLOAD

1 min flash movie with Head motion X graph

Day 3	4 persons with FACE trace

1 min flash movie with Conversation Chart

Day 3	4 persons with FACE trace

Video & Audio data LIST '07/11/05 – '07/11/07

Our latest data is very multi-source (and multi-modal)

Multiple views (and recordings) of the same scene

Conversation Data in English

Video & Audio data LIST '07/11/05 – '07/11/07

注) AVIファイルはクリックしても自動再生はされません。右クリックで保存してからご使用ください。

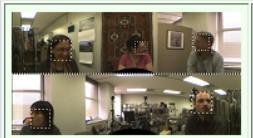
DAY 1 '07/11/05

360 degree face trace
34:35 460MB
download
(AVI)

Angle 1
43:25 444MB
download
(MPG)

1:19:30 187MB
MP3 sound data
download

DAY 2 '07/11/06



360 degree face trace
1:01:16 181MB
download
(AVI)

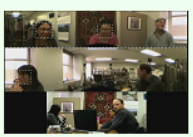
Angle 1
1:22:23 842MB
download
(MPG)

Angle 2
1:22:12 840MB
download
(MPG)

Angle 3
1:11:35 732MB
download
※bad sound(MPG)

SANYO
1:52:43 1,151MB
download
(MPG)

Angle1 + Angle2



360degree+SANYO
1:01:16 626MB

1:48:22 255MB
MP3 sound data
download

Subtitles and multi-view composite video

Angle1 + Angle2
1:22:15 841MB
download
(MPG)

360degree+SANYO
1:01:16 626MB
download
(MPG)

1:48:22 255MB
MP3 sound data
download

DAY 3 '07/11/07

360 degree face trace
1:01:16 181MB
download
(AVI)

Angle 1
1:22:44 846MB
download
(MPG)

Angle 2
1:22:18 841MB
download
(MPG)

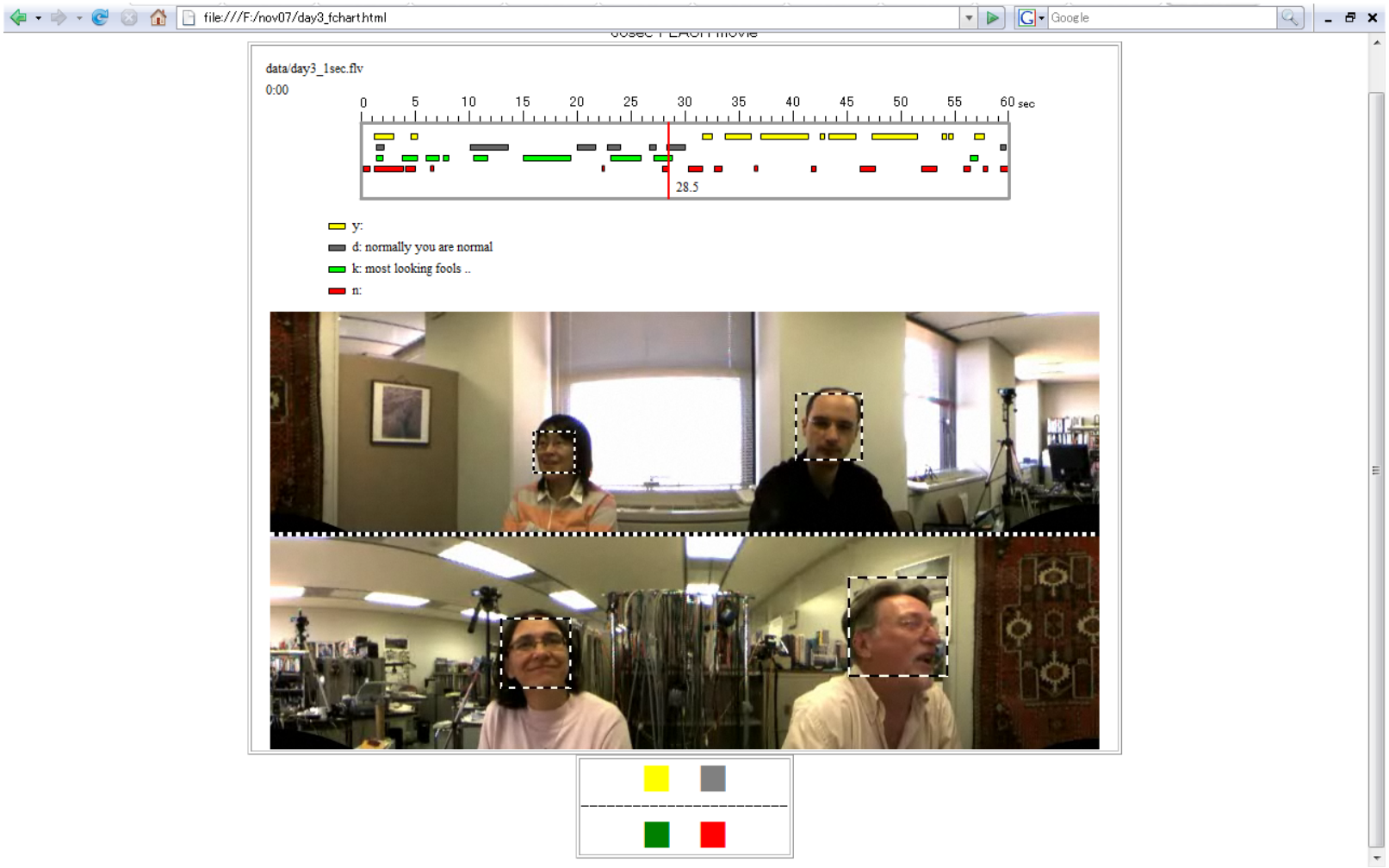
Angle 3
1:03:50 652MB
download
※bad sound(MPG)

360 degree
1:22:43 845MB
download
※bad sound(MPG)

Angle1 + Angle2
1:22:44 1,178MB
download
(MPG)

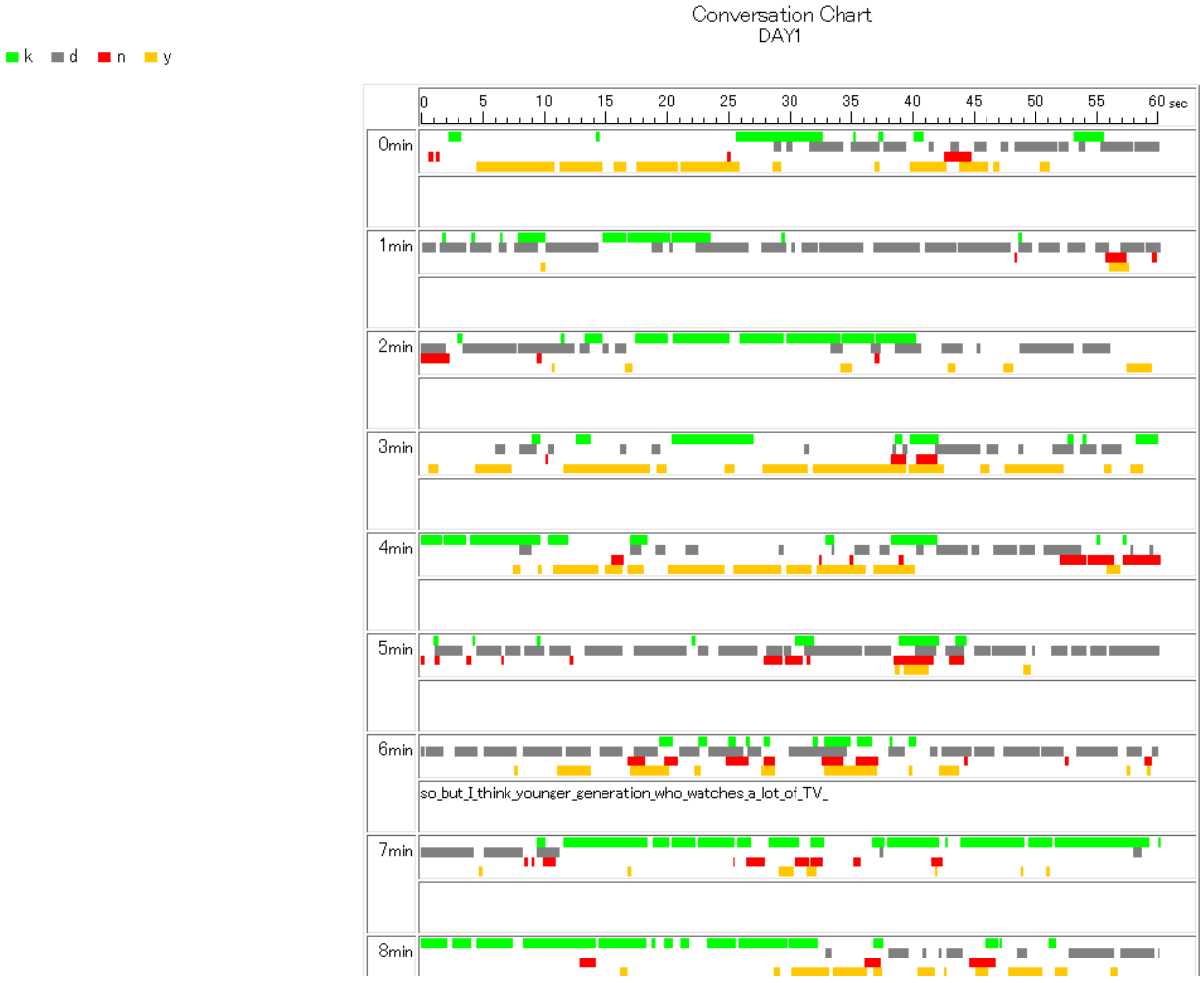
1:33:40 220MB
MP3 sound data
download

Transcriptions scroll as the video plays



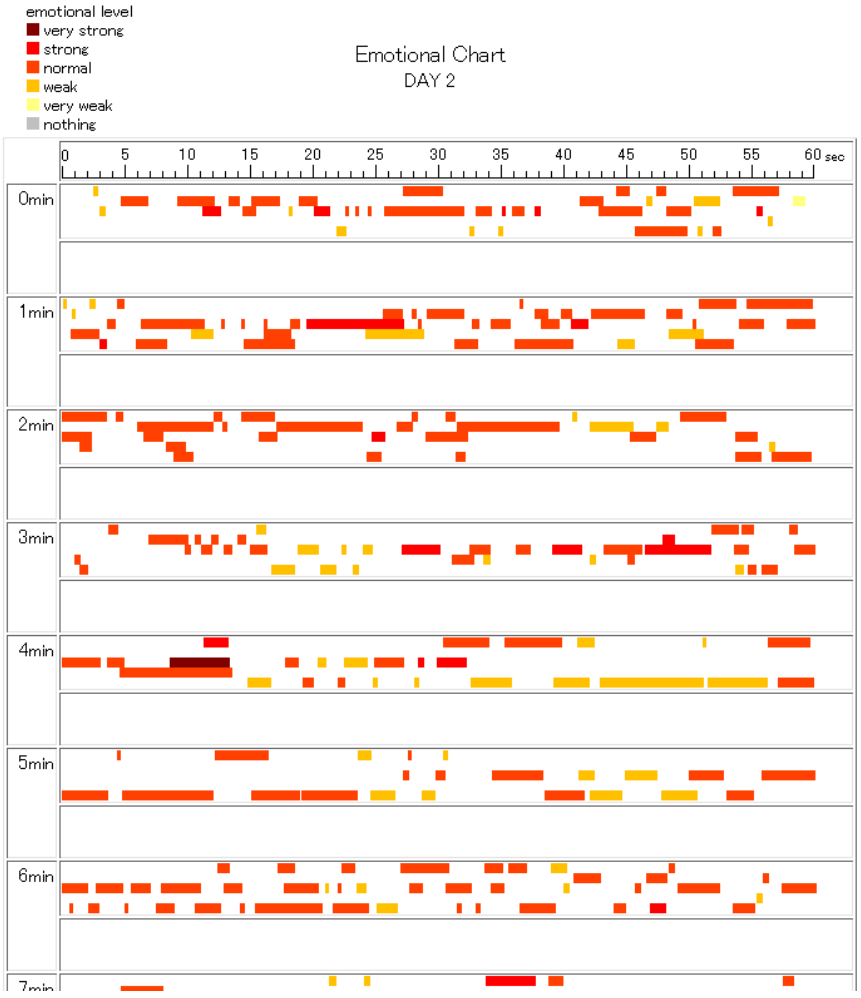
Speech overlaps can be interactively visualised in this way

Scanning Discourse Segments



bar-plot displays can be scrolled with the video (karaoke)

Rapport / Discourse Intensity



Subjective (and automatic) measures of discourse activity

Animated data streaming (synced with the video)

Automatically-determined measures of head/body activity (x,y,z)

topic labels

topic number	topic title	video file counter	what happened at the change of topic	who's mainly talking	who's listening/reacting	mood: heated/ quiet
10	Manzai	0:00:00	Izumi starts talking about how different the said topic "Rakugo" and "Manzai" are.	Izumi	Christina interested and Damien explains his image of Manzai. Nick almost just listening.	interested and heated
20	"Oubeika"	0:01:23	Damien starts picking up from the former topic.	Damien explains the meaning of "Obeika" how it is used. Izumi explains the technique of Manzai that makes the expression "Obeika" effective in Manzai as well as its variation.	Damien actively asks questions and Christina gives her comprehension. Nick seems very interested.	interested and heated
30	Damien used "Obeika" in kyoto	0:05:06	Damien explains what happened in Kyoto.	Damien said "Obeika" on the street.	Everybody seems very interested and laughed a lot.	very funny
40	to get to know fragement of cultual backgraound	0:06:44	Damien draws the topic to general understading.	Damien and Christina	Nick and Izumi mainly listen a bit quietly and later gives their understanding.	a bit quiet
50	"Sonnano kankei nee"	0:08:48	Izumi starts talking about another popular phrase "Sonnano kankei nee"	Izumi explains with the original action as well as the comedian's funny looks in a swimsuit.	Nick also knows a bit about him. Damian wants to know how the expression is used in ordinary Japanese. Christina also interested.	heated but doesn't last long
60	changing the subject	0:10:26	Nick			
70	TV program NHK	0:10:33	Izumi says that the only TV chnnel she could watch in Indonesia was NHK.	Izumi		a bit quiet

topic labels for day-2

topic number	topic	audio file counter	what happened at the change of topic	who's mainly talking	who's listening/ reacting	mood: heated/ quiet	information exchanged: large/small
10	preparation	0:17:13	Izumi and Lorene talked to each other… just a little.	Nick gives instruction.	The others make sure what to do.	busy	small
20	introducting each other (except Christina and Damian)	0:20:33	Damian gives cue and Nick started.	Nick leads intruduction.	Not much informaiton exchanged.	friendly	small
30	Nick explaining to Lorene what happened yesterday and the purpose of this session	0:20:57	Nick started.	Nick	Izumi and Lorene seem interested.	heated at first but unable to extend then quiet	medium
40	Lorene's plan (length of stay and what she has been doing)	0:21:36	Nick started.	Nick	Lorene answers and Christina asks questions.	friendly	small
50	Lorene's work (contract nursing)	0:22:19	Lorene includes in her answer to Christina.	Lorene	Nick and Christina seem inerested.	friendly	medium
60	hospital	0:22:53	extending from Lorene's work	Nick	Lorene answers questions.	friendly	medium
70	Nick's experience (first aid for earthquake)	0:23:13	extending from Lorene's work	Nick	Lorene adds more information and the others seem interested.	interested and heated	large
80	international color coding	0:23:46	Christina asks.	Nick	Lorene adds more information and the others seem interested.	interested and heated	large
90	Walkie-talkie	0:24:04	Nick started.	Nick	Everybody seems interested.	Nick made it fun.	medium
100	satellite	0:24:40	Nick started.	Nick	Everybody seems interested.	Nick made it fun.	small
110	life saving in Australia	0:24:59	Lorene said Izumi has a life saver ticket in Aus.	Lorene	Everybody seems interested.	friendly	medium
120	Izumi's cell phone rang	0:25:45		Nick	Everybody laughs at Nick's joke.	quiet then funny atmosphere	none
130	how each other's name came from	0:25:58	Starting from Lorene talking to Damian	Damian, Izumi, Lorene explaining	Everybody seems interested and asks and talks.	interesting and heated	large
140	Lorene and Izumi went to Tanzan Jinja and enjoyed hotspring.	0:28:33	Lorene changed the topic.	Lorene and Izumi	Nick and Christina seem inerested.	friendly at first but unable to extend	medium

topic labels for day-3

topic number	topic title	video file counter	what happened at the change of topic	who's mainly talking	who's listening/ reacting	mood: heated/ quiet	information exchanged: large /small
10	preparation for recording	0:00:01	Nick is setting for recording	Nick	Damien minds the noise for recording and asks Nick if he can do something. Izumi and Christine are waiting..	a bit quiet	small
20	start recording	0:01:30	Nick is apologizing in Japanese.	Nick	Every one.	little bit heated	small
30	about Izumi's friend	0:02:08	Nick asks about Izumi's friend.	Izumi is explaining what her friend is doing.	Nick is interested in the massage chair and massage Izumi's friend doing. Damien is listening.	little bit heated	small
40	last day	0:03:27	Nick is saying that it is the last day to record.	Nick	Izumi pretends being sad and sobbing.	little bit heated	small
50	feminine Japanese	0:03:35	Nick reacts like Japanese.	Izumi points out that Nick sounds like women and makes fun of it.	Damien says it's typical and often happens when foreign men going out with Japanese women.	little bit heated	large
60	how to use boku	0:04:14	Christina asks if she could use "Boku".	Christina and Nick.	Christina and Nick are interested in if women use "Boku".	little bit heated	midium
70	sounds like female	0:04:56	Izumi imitates the way of Japanese women nod.	Izumi knows a man who was going out with Indonesian girl sounded very female.	everyone laughs, especially Nick.	heated	large
	how to use	0:05:33	Christina asks how to	Damien is explaining	Nick adds	little bit heated	large

outside

Conclusion

We are working with large quantities of dialogue speech including business meetings, friendly discourse, and telephone conversations, and have produced web-based tools for the visualisation of non-verbal and paralinguistic features of the speech data.
This paper has described some of our tools and techniques for accessing large quantities of speech data and for the visualisation of discourse interactions and events at levels above that of linguistic content.
In essence, they provide higher-level displays so that specific sections of speech, text, or other annotation can be accessed by the researcher and provide an interactive interface to the large amount of data through an Archive Browser.

Conclusion (cont.)

thanks

Thank you for listening

further information

data and examples can be found at

http://feast.atr.jp/nonverbal/