



Rec 2008 marrakech

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!!

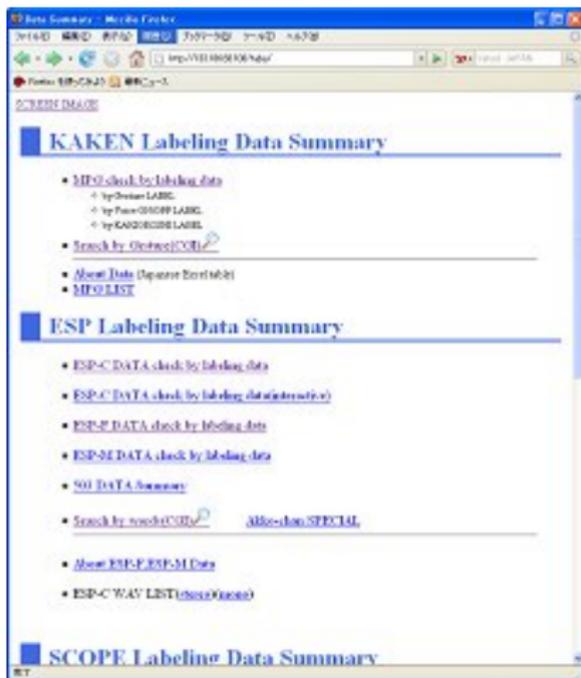
New ways to access the data

- 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 information
- accessed through ordered search
- new annotations linked as data grows and is shared

<http://feast.atr.jp> : top



The screenshot shows a web browser window with the URL <http://feast.atr.jp>. The page title is "KAKEN Labeling Data Summary". The content includes a list of links for KAKEN data, followed by a section for "ESP Labeling Data Summary" with several links, and a partially visible "SCOPE Labeling Data Summary" section at the bottom.

KAKEN Labeling Data Summary

- [MFL check by labeling data](#)
 - by [Stream LABEL](#)
 - by [Para-GRAPH LABEL](#)
 - by [KASUGI/CD LABEL](#)
- [Search by \[Gotohji/CD\]\(#\)](#)
- [About Data Separator Error Model](#)
- [MFL LIST](#)

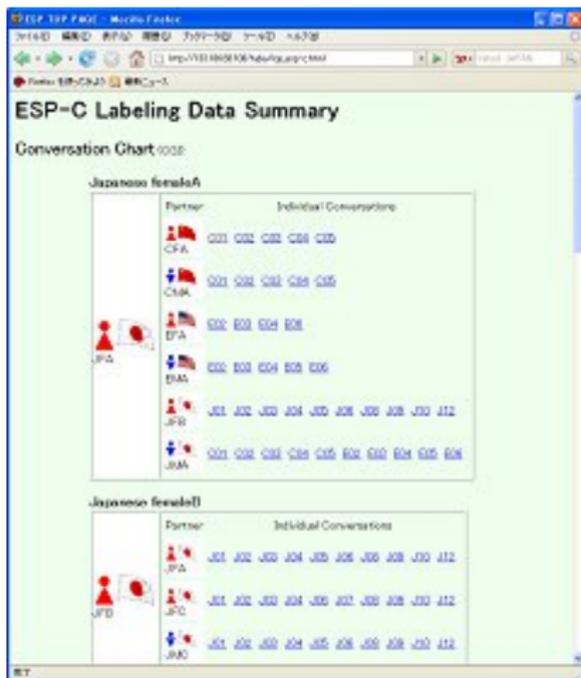
ESP Labeling Data Summary

- [ESP-C DATA check by labeling data](#)
- [ESP-C DATA check by labeling data \(structure\)](#)
- [ESP-F DATA check by labeling data](#)
- [ESP-SE DATA check by labeling data](#)
- [S01 DATA Summary](#)
- [Search by \[nochi/CD\]\(#\)](#) [Jike-shou/SPECTAL](#)
- [About ESP-F, ESP-M Data](#)
- [ESP-C WAV LIST \(Index/Kakasi\)](#)

SCOPE Labeling Data Summary

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)

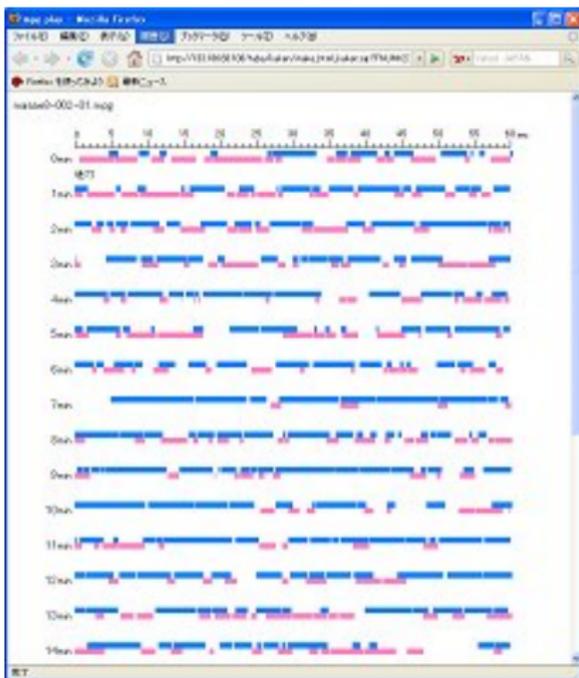
KAKEN Labeling Data Summary

• MPG Check by labeling data (005)

	MPG Name	contents	min	Labeling Data			KAKENID
[All MPG]	nhaku0-001-01	お笑い、音楽、U.S.D.S.	42min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku0-002-01	お笑い、音楽、U.S.D.S.	37min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku1-001-01	お笑いA	28min	GESTURE	IDPR	VOICE on/off	SAKEKOSHI (1min)
[All MPG]	nhaku1-002-01	お笑いA	42min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku2-001-01	お笑いA	38min	GESTURE	IDPR	VOICE on/off	SAKEKOSHI (1min)
[All MPG]	nhaku2-002-01	お笑いA、音楽	14min	GESTURE	IDPR	VOICE on/off	SAKEKOSHI (1min)
[All MPG]	nhaku2-004-01	お笑いA	11min	GESTURE	IDPR	VOICE on/off	SAKEKOSHI (4min)
[All MPG]	nhaku3-001-01	お笑いA	51min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku3-002-01	お笑いA	21min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku4-001-01	お笑いA	40min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku4-002-01	お笑いA	42min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku5-001-01	お笑い、音楽、U.S.D.S.	37min	GESTURE	IDPR	VOICE on/off	
[All MPG]	nhaku6-000-01	お笑い、音楽	77min	GESTURE	IDPR	VOICE	

Multimodal data - annotated video & audio

Visualising a conversation



Speech activity for the two participants, showing overlaps & turns

View of a Multi-Party Dialogue

Conversation (BT - Multi-Party)

http://193.100.106.146/lego/lego/analysis/analysis.html

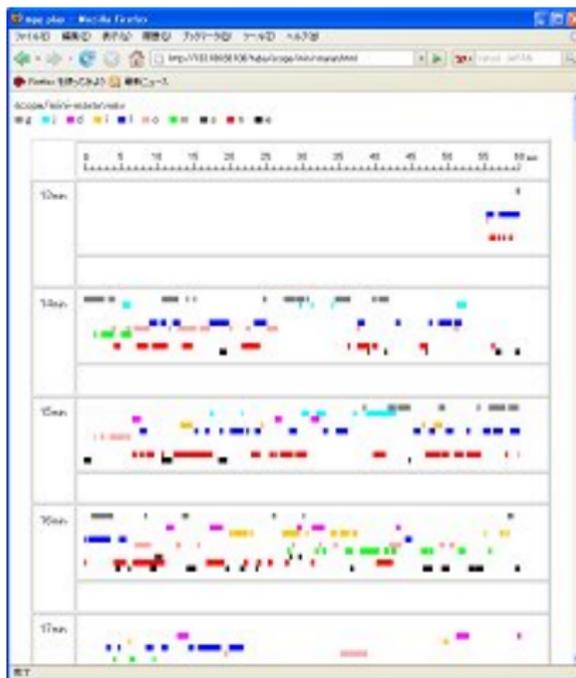
FileView 印刷/印刷範囲 検索Ctrl-F

FAN001_00 Conversation List

start	end	speaker	PLAY	color	in	to	text
1000	6.75	FAN001_00	PLAY	赤	赤	赤	見物の説明
10.301	10.750	FAN001_00	PLAY	赤	赤	赤	このままではだめだ
11.102	11.820	FAN001_00	PLAY	赤	赤	赤	ふん
12.043	12.810	FAN001_00	PLAY	赤	赤	赤	そう
12.405	12.250	FAN001_00	PLAY	赤	赤	赤	この中にアリス
16.261	14.200	FAN001_00	PLAY	赤	赤	赤	これで最後じゃない
14.433	16.40	FAN001_00	PLAY	赤	赤	赤	くまのこやうしてどう
16.710	19.200	FAN001_00	PLAY	赤	赤	赤	一週間(10)程度(10)
16.962	16.000	FAN001_00	PLAY	赤	赤	赤	ふん
20.500	21.700	FAN001_00	PLAY	赤	赤	赤	それだけだ
21.940	22.740	FAN001_00	PLAY	赤	赤	赤	それだけ
23.910	24.900	FAN001_00	PLAY	赤	赤	赤	それだけじゃ、だめだ
25.440	27.200	FAN001_00	PLAY	赤	赤	赤	それだけじゃ、だめだ
28.112	28.730	FAN001_00	PLAY	赤	赤	赤	くまのこやうして(10)
28.311	30.130	FAN001_00	PLAY	赤	赤	赤	くまのこやうして(10)
31.261	32.190	FAN001_00	PLAY	赤	赤	赤	そう
32.141	33.250	FAN001_00	PLAY	赤	赤	赤	日本人だ
33.254	33.000	FAN001_00	PLAY	赤	赤	赤	日本人だ
34.261	34.600	FAN001_00	PLAY	赤	赤	赤	日本人だ
34.881	37.000	FAN001_00	PLAY	赤	赤	赤	日本人
37.000	38.30	FAN001_00	PLAY	赤	赤	赤	くまのこやうして(10)
38.400	39.000	FAN001_00	PLAY	赤	赤	赤	くまのこやうして(10)
41.000	43.30	FAN001_00	PLAY	赤	赤	赤	くまのこやうして(10)

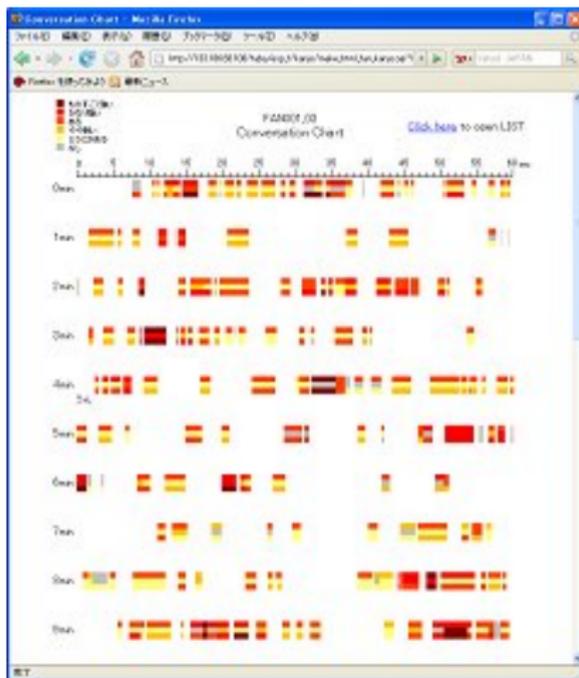
In a multi-party discourse, each speaker is colour-coded

Display of a multi-person dialogue



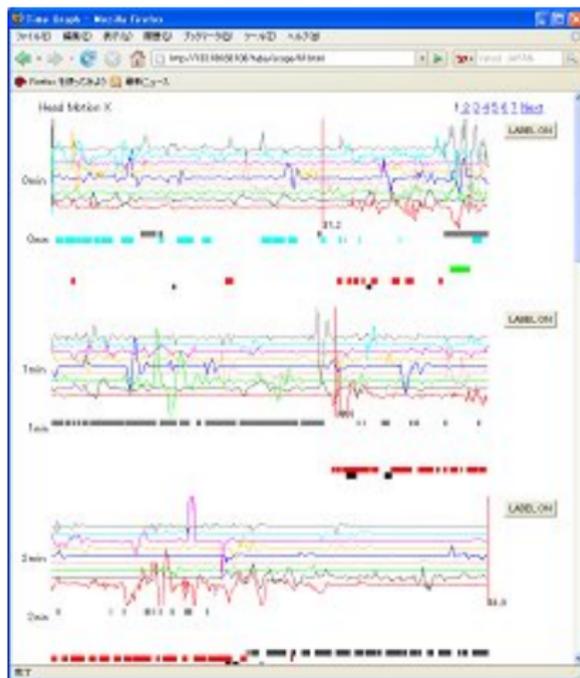
Colour-coded bar-plots allow a better view of the interaction

Labelling the “intensity” of a discourse (rapport)



Three labellers indicate discourse involvement for 1 speaker

Plotting head and body motion



Or movement data plotted using the same colour-coding

Gesture

KAKEN Search Gesture Label (○○○)

Select GESTURE:
両手
片手
両足
片足

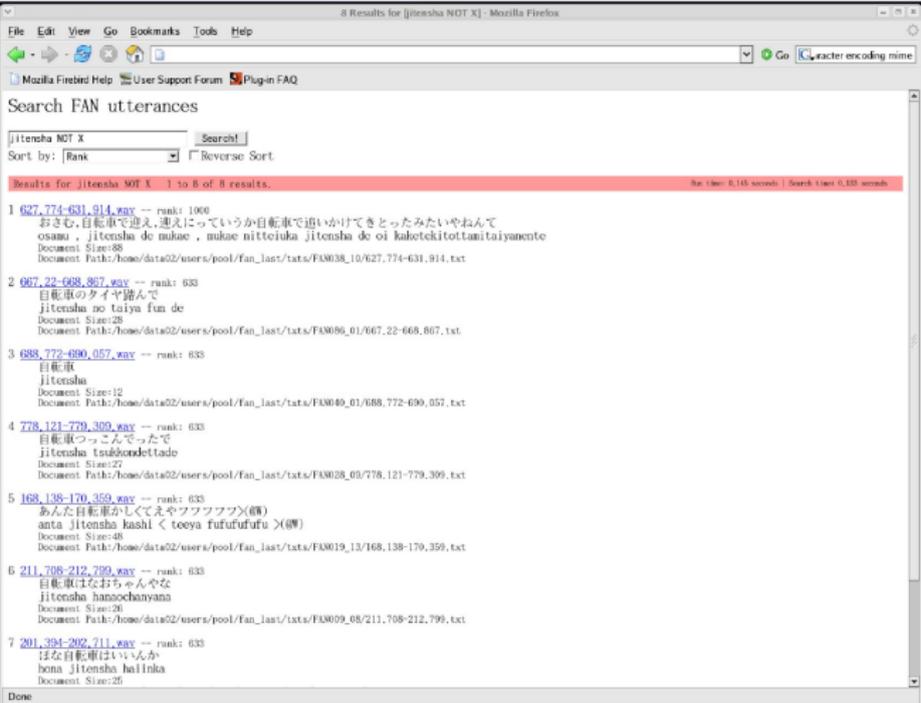
検索 START

Select FILE:

Search?	MPQ filename	contents	file
<input type="checkbox"/>	nakav-001-01	全知人、作業、挨拶	43min
<input type="checkbox"/>	nakav-002-01	全知人、作業、挨拶	22min
<input type="checkbox"/>	nakav-001-01	全知人	28min
<input type="checkbox"/>	nakav-002-01	全知人	43min
<input type="checkbox"/>	nakav-001-01	全知人	28min
<input type="checkbox"/>	nakav-002-01	全知人、手紙	14min
<input type="checkbox"/>	nakav-004-01	全知人	11min
<input type="checkbox"/>	nakav-001-01	全知人	67min
<input type="checkbox"/>	nakav-002-01	全知人	27min
<input type="checkbox"/>	nakav-001-01	全知人	46min
<input type="checkbox"/>	nakav-002-01	全知人	42min
<input type="checkbox"/>	nakav-001-01	全知人、作業、挨拶	37min
<input type="checkbox"/>	nakav-002-01	全知人、作業、挨拶	22min
<input type="checkbox"/>	nakav-001-01	全知人	28min
<input type="checkbox"/>	nakav-002-01	全知人電話会話	50min

individual examples of a given gesture can be listed for viewing

SWISH - a Google-type interface to large text data



The screenshot shows a Mozilla Firefox browser window with the title "8 Results for [jitensha NOT X] - Mozilla Firefox". The address bar contains "http://www.mozilla.org". The page content is titled "Search FAN utterances" and features a search box with the text "jitensha NOT X" and a "Search!" button. Below the search box, it says "Sort by: |Rank" and "Reverse Sort". A red banner indicates "Results for 'jitensha NOT X' 1 to 8 of 8 results." with a search time of 0.145 seconds. The results are numbered 1 through 7, each showing a file name, rank, a snippet of Japanese text, document size, and file path.

Rank	File Name	Rank	Snippet	Document Size	Document Path
1	627-774-631_914.wav	1060	おさむ、自転車で迎え、迎えについてどうか自転車で追いかけてきとったみたいやおんて osamu, jitensha de mukae, mukae nitteiuika jitensha de oi kakotekittotamitaiyante	98	home/data02/users/pool/fan_last/tats/FAM038_10/627-774-631_914.txt
2	667-22-668-867.wav	633	自転車のクイヤ諾んで jitensha no taiya fun de	28	home/data02/users/pool/fan_last/tats/FAM086_01/667-22-668-867.txt
3	688-772-690_057.wav	633	自転車 jitensha	12	home/data02/users/pool/fan_last/tats/FAM049_01/688-772-690_057.txt
4	778-121-779-309.wav	633	自転車つっこんでったで jitensha tsukkonde tade	27	home/data02/users/pool/fan_last/tats/FAM028_00/778-121-779-309.txt
5	168-138-170_359.wav	633	あんだ自転車かしくてえやフフフフフ>(H) anda jitensha kashiki toeya fufufufufu >(H)	48	home/data02/users/pool/fan_last/tats/FAM019_13/168-138-170_359.txt
6	211-708-212-799.wav	633	自転車はなおちゃんやな jitensha hanoochanyana	26	home/data02/users/pool/fan_last/tats/FAM009_08/211-708-212-799.txt
7	201-394-202_711.wav	633	ほな自転車はいんか hona jitensha ha'inika	25	

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

MFG data with subtitles '07/11/05 - '07/11/07



Conversation Chart & LIST

data	length		CHART	Emotion CHART	LIST	TOPIC LIST	MFG data
DAY 1	34:05	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 '07/11/06 - '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 '07/11/06



360 degree face trace
34:35 480MB
[download](#)
(AVI)



Angle 1
43:25 444MB
[download](#)
(MPG)

1:19:30 187MB
MPG 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
MPG 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
MPG 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)



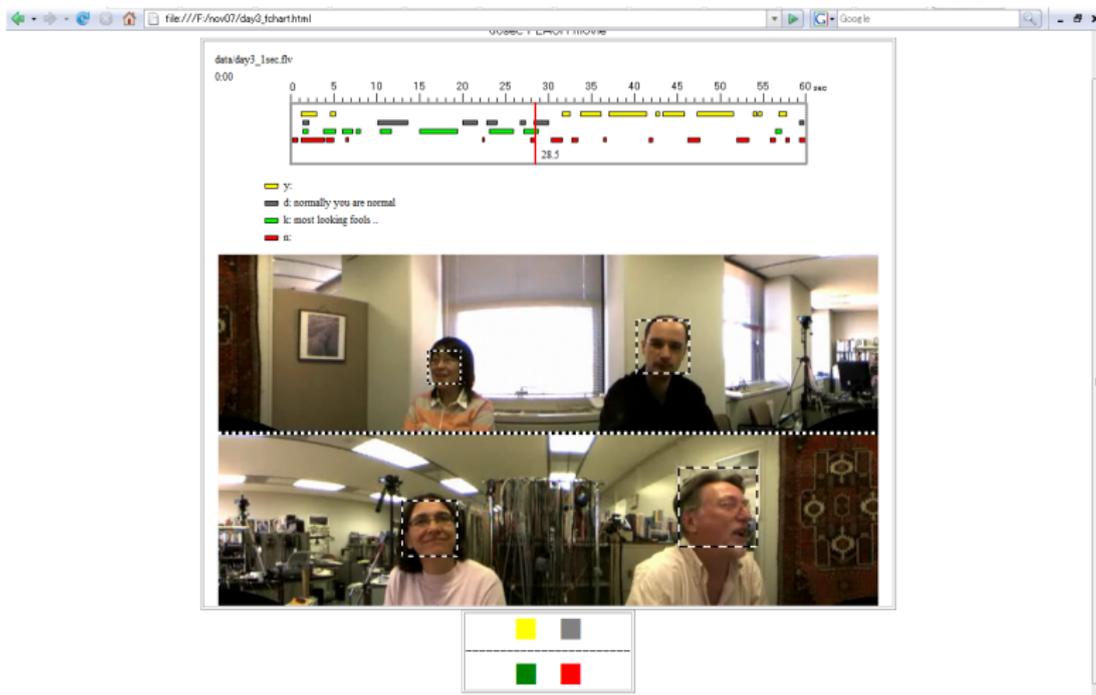
360 degree
1:22:43 845MB
[download](#)
❌bad sound(MPG)



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

1:33:40 220MB
MPG sound data
[download](#)

Transcriptions scroll as the video plays



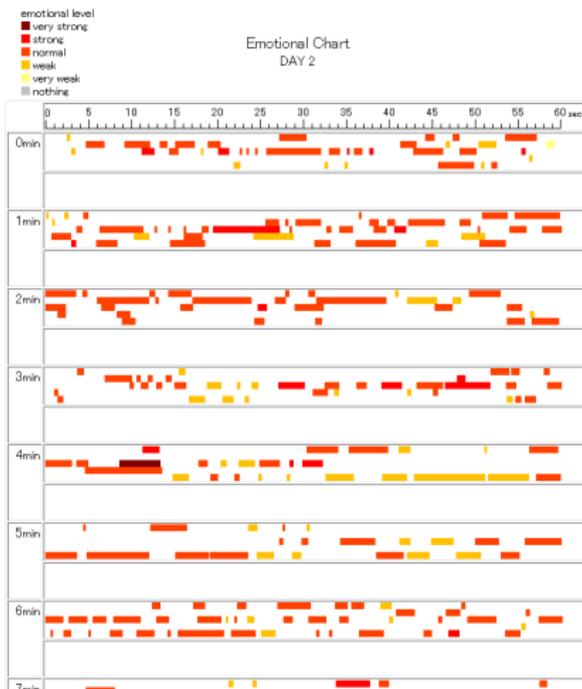
Speech overlaps can be interactively visualised in this way

Scanning Discourse Segments



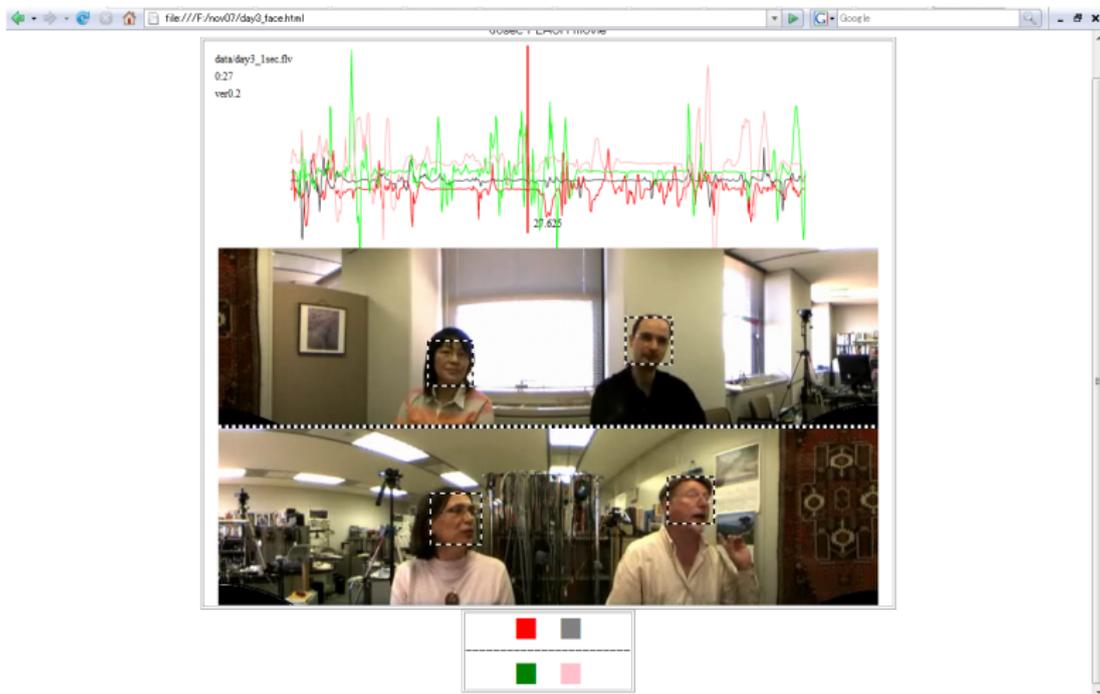
bar-plot displays can be scrolled with the video (karaoke)

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
		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
10	Manzai	0:01:23	Damien starts picking up from the former topic.	Damien explains the meaning of "Obelika" how it is used. Izumi explains the technique of Manzai that makes the expression "Obelika" effective in Manzai as well as its variation.	Damien actively asks questions and Christina gives her comprehension. Nick seems very interested.	interested and heated
20	"Obelika"	0:05:06	Damien explains what happened in Kyoto.	Damien said "Obelika" on the street	Everybody seems very interested and laughed a lot.	very funny
30	Damien used "Obelika" in kyoto	0:06:44	Damien draws the topic to general understanding.	Damien and Christina	Nick and Izumi mainly listen a bit quietly and later gives their understanding.	a bit quiet
	to get to know fragment of 40 cultural background	0:08:48	Izumi starts talking about another popular phrase "Sonnano kankai 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. Damien wants to know how the expression is used in ordinary Japanese. Christina also interested.	heated but doesn't last long
50	"Sonnano kankai nee"	0:10:26	Nick			
60	changing the subject	0:10:33	Izumi says that the only TV channel she could watch in Indonesia was NHK.	Izumi		a bit quiet
70	TV channel NHK					

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	01713	Izumi and Lorene talked to each other"" Just a little.	Nick gives instruction.	The others make sure what to do.	busy	small
20	introducing each other (except Christina and Damian)	02033	Damian gives cue and Nick started.	Nick leads introduction.	Not much information exchanged.	friendly	small
30	Nick explaining to Lorene what happened yesterday and the purpose of this session	02057	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)	02136	Nick started.	Nick	Lorene answers and Christina asks questions.	friendly	small
50	Lorene's work (contract nursing)	02219	Lorene includes in her answer to Christina.	Lorene	Nick and Christina seem interested.	friendly	medium
60	hospital	02253	extending from Lorene's work	Nick	Lorene answers questions.	friendly	medium
70	Nick's experience (first aid for earthquake)	02313	extending from Lorene's work	Nick	Lorene adds more information and the others seem interested.	interested and heated	large
80	international color coding	02346	Christina asks.	Nick	Lorene adds more information and the others seem interested.	interested and heated	large
90	Walkie-talkie	02404	Nick started.	Nick	Everybody seems interested.	Nick made it fun.	medium
100	satellite	02440	Nick started.	Nick	Everybody seems interested.	Nick made it fun.	small
110	life saving in Australia	02459	Lorene said Izumi has a life saver ticket in Aus.	Lorene	Everybody seems interested.	friendly	medium
120	Izumi's cell phone rang	02545		Nick	Everybody laughs at Nick's joke.	quiet then funny atmosphere	none
130	how each other's name came from	02558	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 Tanzania Inja and enjoyed hot spring.	02833	Lorene changed the topic.	Lorene and Izumi	Nick and Christina seem interested.	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
	preparation for 10 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
	about Izumi's 30 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
	feminine 50 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
	how to use 60 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
	sounds like 70 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 use	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/>