GRIPS Discussion Paper 24-11

Comparing Verbal and Non-verbal Emotions in Parliamentary Speeches

By

MASUYAMA Mikitaka

September 2024

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 106-8677

Comparing Verbal and Non-verbal Emotions in Parliamentary Speeches

MASUYAMA Mikitaka

National Graduate Institute for Policy Studies

Summary

The Japanese Diet uses the Automatic Speech Recognition system, which directly transcribes parliamentary speech in both plenary and committee meetings. Speech recognition performance is monitored for most meetings, consistently achieving over 90 percent accuracy in Japanese characters. Using the ASR for Diet deliberations, we have developed an internet video retrieval system to create timestamp data to match the minutes of parliamentary meetings and video feeds. Our "Video Retrieval System for Diet Deliberations" allows one to pinpoint and play the parliamentary video clips corresponding to the meeting minutes by keyword search. In this paper, we offer an overview of the video retrieval system and suggest various ways we can utilize our video retrieval system. By exploring how our video retrieval system can generate data to compare verbal and non-verbal emotions, we depart from the tradition of focusing on the minutes and shed new light on parliamentary discussion's complex and multifaceted nature.

Keywords: Information technology, Speech recognition, Video retrieval, Keyword search, Parliamentary discussion, Big Data, Communication, Democracy, Institutions, Internet, Parliaments, Social Media, Technology

This work is supported by JSPS Kakenhi Grant Numbers 15H05727 and 20H00062.
The paper is prepared for presentation at the 2024 American Political Science Association Annual Meeting, Philadelphia, September 5 - 8, 2024.

Introduction

There has been a surge in the development of analytical tools and techniques for analyzing the textual data of parliamentary proceedings. However, with the growing trend of parliamentary video streaming, there is a pressing need for similar tools to analyze audio-visual data. While visual data offers a clear advantage over textual data for a more comprehensive analysis of parliamentary discussions and debates, it can be challenging to pinpoint the exact scene of a particular utterance by a specific speaker in lengthy video recordings that can span for hours.

We have launched an internet video retrieval system for the Japanese Diet to remedy such a situation. Using the latest speech recognition techniques to create timestamp data to match parliamentary video feeds and the minutes of proceedings, it can pinpoint and play the parliamentary video clips corresponding to the minutes of proceedings through keyword search. With our video retrieval system, one can directly retrieve the video feed segment one is particularly interested in, understand the flow of parliamentary debates, and check the speaker's facial expressions and body language. In addition, our system captions the videos, offering an alternative means of accessing parliamentary deliberation clips for those with hearing impairments. Since it is easy to share the URL identifying a moment in a video feed via SNS, our system has great potential to boost the usage of Diet deliberation videos by researchers and ordinary citizens.

This paper offers an overview of the video retrieval system we have developed and demonstrates how one can retrieve video streaming on user terminals that do not support Japanese language input. We also suggest various ways to utilize our video retrieval system and explore how our video retrieval system can generate data to compare verbal emotions in speeches and non-verbal emotions in facial expressions.

Video Retrieval System for Diet Deliberations

The Japanese Constitution stipulates that each House of the Diet shall keep a record of proceedings and put it into general circulation. Since the opening of the Imperial Diet in 1890, verbatim records have been made by manual shorthand. However, early in this century, the House of Representatives, one of the two Houses of the Diet, terminated recruiting stenographers and investigated alternative methods for transcribing parliamentary speeches.

Kawahara, one of the research partners, has developed Automatic Speech Recognition (ASR) technology, which has been deployed in the transcription system for the House of Representatives. To achieve high recognition performance in spontaneous meeting speech, Kawahara has investigated an efficient training scheme with minimal supervision that can exploit a large amount of actual data and proposed a lightly supervised training scheme based on statistical language model transformation, which fills the gap between faithful transcripts of spoken utterances and final texts for documentation. Once the mapping is trained, faithful transcripts for training acoustic and language models are no longer needed. The ASR system has consistently achieved character accuracy of over 90% since 2011, which helps streamline the transcription process. The accuracy rate has currently improved to 95 percent.[1]

The Diet Library currently provides the digitized minutes of parliamentary meetings via the Internet. On the other hand, we can watch the live online streaming of

[1] Kawahara (2012 and 2024).

proceedings at the secretariat website of each house. We can also search the video library and watch video streaming of parliamentary proceedings. Both houses originally made video streaming of plenary and committee meetings available for only one year, but the House of Representatives changed its policy so that the videos of proceedings since 2010 are currently available for viewing.

https://www.shugiintv.go.jp/jp/index.php

https://www.webtv.sangiin.go.jp/webtv/index.php

Diet deliberation videos can be searched by meeting date, meeting title, subject, and speaker, although only the first two search options are available in the English interface. However, even if we successfully retrieve the desired Diet deliberation video, we must watch the video streaming from the beginning to the speech or debate segment we are particularly interested in. It is not uncommon for a committee meeting to last more than 7 hours. While the video breakdown by questioner is available in the Japanese interface, streamed video segments are usually 30 to 60 minutes long. No such breakdown is available in the English interface. Moreover, replies to parliamentary questions are included in the video, arranged by the questioner. Thus, we cannot search the Diet deliberation videos of prime and cabinet ministers answering parliamentary questions.

By linking the Diet Library's proceedings database and the Diet secretariats' deliberation video libraries, our "Video Retrieval System for Diet Deliberations (VRS)" makes it possible to retrieve the deliberation video clips corresponding to the minutes of the proceedings through keyword searching.[2]

https://gclip1.grips.ac.jp/video/

Unlike the Diet secretariats' websites, our VRS creates and adds subtitles to the Diet deliberation videos, thus offering those with hearing impairment access to the deliberation videos. Without relying on text vocalizing software that produces somewhat synthetic voices that do not resemble that of the original speaker, our VRS provides better access to instantly listening to what was actually spoken in the Diet for those with optical impairment.

Technically speaking, our VRS consists of two sub-systems. As illustrated in Figure 1, one of the sub-systems uses the latest speech recognition techniques to create timestamp data to match the Diet Library's proceedings database (Minute DB) and the Diet secretariats' deliberation video databases (Video DB). The second sub-system uses the timestamp data to search the Diet proceedings and retrieve the Diet deliberation videos corresponding to the minutes by keyword search (Web-based Search Interface). The results of keyword searches are deliberation video links, and the portion of the video we are particularly interested in can be played partially by clicking the URL link for the deliberation video stored in the Diet secretariats' databases (not stored in our VRS).

Our VRS has been in operation and publicly available since November 2012. It is possible to keyword search all the plenary and committee meetings in the House of Representatives since January 2010 and those in the House of Councillors since December 2012. Below, we briefly describe how our VRS works. Figure 2 shows the top page of our web-based search interface, allowing us to search for deliberation video segments by typing keywords. The Japanese interface will appear when the user clicks "Japanese" in the upper right-hand corner.

[2] Masuyama and Kawahara (2019) and Masuyama et al. (2024).

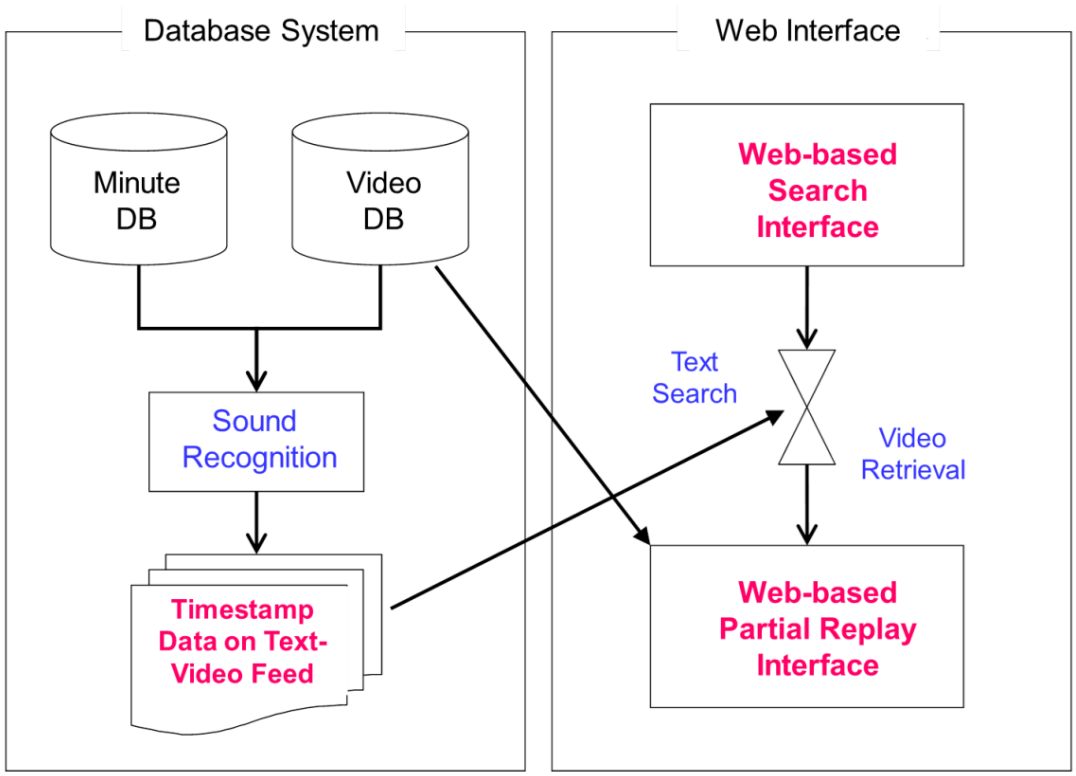


Figure 1: Process for Linking Diet Proceedings and Deliberation Videos

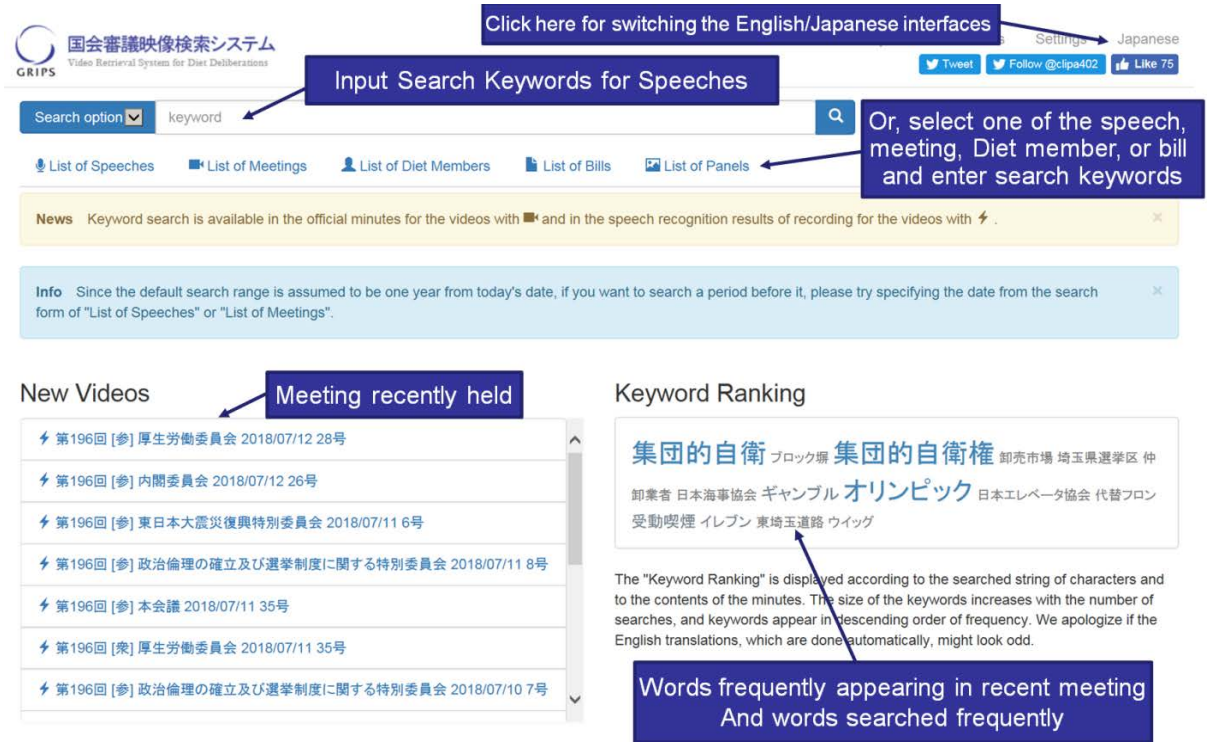


Figure 2: Keyword Search Interface in English

One can type English keywords separated by spaces in the search field, and they will be translated automatically into Japanese and used in keyword searching. Any combination of keywords is acceptable. For instance, if one types "Kishida Fumio" (the name of the current Prime Minister of Japan) and "tax increase" in the search field and hits the search button, a list of the search results will appear in ascending order of date (Figure 3). Alternatively, one could first select one of the five categories (speech, meeting, Diet member, bill, and panel) from the "search option" and type keywords in the search field. As the default setting, our system searches the database for the past year, although it can be extended or shortened by calendar and filtered by other factors in the search results interface. Then, one can click one of the video links, and our system will instantly play the portion of the video corresponding to the speech, including the keywords.

Figure 3: Video Feed Link

In addition to specifying AND/OR search options by selecting one of the boxes under the search field, we can filter the search results by date and limit the search to one or both of the two houses or joint meetings of the two houses. In the filtering area, additional search options list speakers (top 5) and meetings (top 10) with words with utterances matching the search text, allowing us to narrow the search results by selecting one of the meetings and one of the speakers.

If we click one of the video feed links, the screen content similar to that in Figure 4 will appear. Subtitles are shown in the caption area. The speech list is shown on the

right side, and the speech currently playing is highlighted. The video will play for either one minute or three speeches. Alternatively, we can keep playing the video by clicking the play button in the toolbar under the video. By double-clicking any speech in the list on the right side, we can instantly watch the video stream of the speeches before and after the speech found by keyword search. Once the user has moved on to another speech, the original speech found by keyword search remains highlighted in yellow.

Figure 4: Video Replay Interface

Further, the URL for the corresponding video streaming segment is shown below the video, and we can easily share the URL via SNS by clicking the tweet button while the video stream is playing. The text of the speech and the URL will immediately appear in the tweet box after the tweet button is clicked. At the bottom of the page, the profile of the speaker is provided, followed by a list of agendas and a list of the Diet members attending the meeting (not shown in Figure 4)

To assist keyword searching, "Keyword Ranking" on the right side of the top page, as seen in Figure 2, lists 15 words uttered in the Diet proceedings, in descending order of frequency, placing more weight on frequency in later parliamentary meetings than earlier and more weight on single meetings than multiple meetings. The font size of the words increases with the number of searches, reflecting the attention given to Diet deliberation videos. An English translation pops up when the mouse hovers over any word.

We can use our VRS in a variety of ways. For instance, we can create a search results list with the query "tax increase" and the speaker's name, "Kishida Fumio." By clicking one of the video links in the list, we can instantly retrieve a video of Prime Minister Kishida Fumio's speeches, with his voice, facial expressions, and body language, in which he mentioned a tax increase.

Diet members are increasingly posting information about their activities on the websites. Some use their websites to display the minutes of parliamentary proceedings, and some even edit and upload deliberation videos on their websites. In contrast, our video retrieval system allows us to obtain the URL for a moment of video streaming and to create a list of video links without downloading and editing the video files.

Furthermore, we can use a tweet function to create a list of parliamentary speeches. For instance, ministers customarily begin answering questions in plenary meetings by saying, "There is a question regarding X." Thus, we can narrow the search mentioned above to speeches beginning with "There is a question regarding tax increase," and tweet the speeches and their video links to create a list of Prime Minister Kishida's plenary speeches on the issue of tax increase.

Another way of utilizing the interfaces for keyword searching and partial replay is to post deliberation video links to the internet news. For instance, if there is a report on a newspaper website featuring the remarks made by Prime Minister Kishida in the Diet, we can enhance the internet news visually by using our VRS and inserting the video link for the moment of video streaming in question. Clicking on the link will result in the instant playing of the video of the moment of Prime Minister Kishida's controversial remark.

The minutes of the proceedings are an essential source of the content of discussion in the Diet, but they do not tell the whole story. For instance, legislators often use supplementary materials in parliamentary meetings and refer extensively to graphic materials such as figures and tables. Such supplementary materials are not included in the minutes unless a Diet member attending the meeting requests that the minutes include them. Although the secretariats and the Diet Library keep the supplementary materials used in parliamentary meetings, those materials are not widely known and hardly used.

We have developed pattern recognition techniques to distinguish between the portions of videos that do or do not focus on the speaker and automatically extract video clips that focus on supplementary materials used in committee meetings. Moreover, the minutes are silent regarding non-verbal communication, and we are developing a web-based program to automatically extract and analyze the speaker's facial expressions and body language. Figure 5 illustrates the supplementary materials found by pattern recognition, compiled into the database using text recognition in video images.

A unique aspect of our VRS is that we use speech recognition techniques to create timestamp data to match Diet proceedings and deliberation videos. In other words, we deal with two types of text information related to parliamentary meetings. The information derived from speech recognition is "correct" since it captures 100 percent of what was spoken in the Diet. However, it may contain irrelevant filler and words that are wrongly recognized due to individual speaker factors such as intonation and pronunciation and technical and environmental factors such as recording quality and noise. On the other hand, the minutes of the Diet meetings become "official" after transcription to eliminate filler, correct inappropriate wording, and add commas and periods so that the speech in the Diet can make sense as a written language.

With a web-based program to automatically calculate correspondence rates as part of the standard procedure for creating timestamp data to match proceedings and deliberation videos, we can systematically analyze the correspondence between the official minutes and speech recognition results by meeting, speaker, and so on in cross-sectional and longitudinal manners.

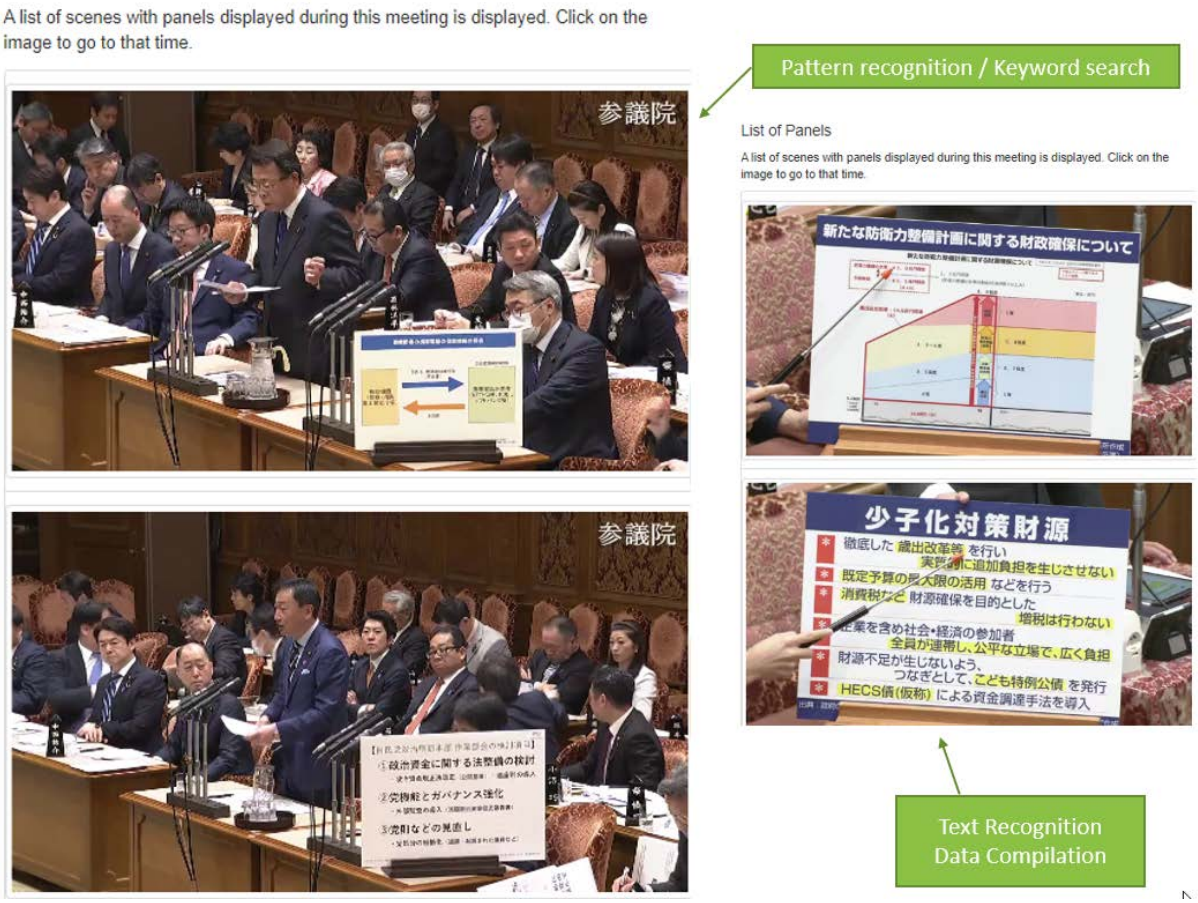


Figure 5: Supplementary Materials Found by Pattern Recognition

Comparing Verbal and Non-verbal Emotions

In this section, we explore the possibility of using our VRS to analyze non-verbal information from the viewpoint that parliaments are organizations that institutionally and systematically accumulate text, audio, and video information. By utilizing our VRS, which uses speech recognition to obtain text data and synchronizes the text data and the time of parliamentary deliberation, it is possible to efficiently check the audio and video information corresponding to the speeches made during parliamentary deliberation.

For example, Japan is the only country in the world whose name has both “Nihon” and “Nippon” pronunciations, and according to an analysis using the Corpus of Spontaneous Japanese by the National Institute for Japanese Language and Linguistics, “Nihon” is overwhelmingly common, and “Nippon” is relatively common in combinations such as “Japan’s No.1” and “Japan’s national team. In our VRS, deliberation videos amenable for keyword search are available from the 174th Diet session that started on January 18, 2010, for the House of Representatives and from the 182nd Diet session that started on December 26, 2012, for the House of Councillors. A search for “Japan” yields 476,994 speeches that include “Japan” and 285,697 in the House of Representatives alone. For example, even when limited to “the Constitution of Japan,” the database includes 3,560 speeches and 2,044 in the House of Representatives. While it is not clear which reading is being made simply by looking at the minutes, our VRS allows one to check reading efficiently by utilizing the functions of keyword searching and instant partial playing of deliberation videos. In linguistics, for example, an attempt has been made to

identify the political stance of American legislators on how they pronounce the name of a foreign place, such as the letter "a" in Iraq as/æ/or/ɑː/.[3] For example, Korean names are read in Hangul or Japanese in Japan, although we cannot check reading in the written records if the names appear in Chinese characters. However, our VRS offers an efficient way to check the reading of Korean names and examine whether the difference in reading has something to do with the partisanship or diplomatic stance of the speakers.[4]

What follows is a preliminary attempt to simultaneously grasp textual and visual information by utilizing the advantages of our VRS, enabling us to pinpoint the video corresponding to a speech and to check the speaker's facial expression, which cannot be done by reading the minutes. To be concrete, we will explain how our VRS can provide the data to compare the verbal emotions in speeches and the non-verbal emotions in facial expressions.

As explained in the previous section, when the deliberation video becomes available on the websites of the parliamentary secretariats, our VRS automatically identifies the URL of the video, obtains the streaming data of the video, extracts the audio data, and applies speech recognition to the audio data. The ARS synchronizes the text resulting from speech recognition with the deliberation video, and our VRS stores the synchronization information between the text and the video, enabling keyword searches and partial play of the deliberation video. When the text data, after going through the transcription, eliminating fillers and misstatements, and becoming publicly available on the Diet Liibrary's website, our system creates text-synchronization information with the deliberation video using the ARS and replaces the synchronization information of the speech-recognition version with it. Through this automatic process, our VRS stores the text records of the parliamentary minutes after a deliberation video becomes available to watch at either of the secretariat websites, which is used for play screen transition by speech and subtitling for each speech and also analyzes the text information through automatic summary and word cloud functions (Figure 6). The latest development of our VRS is a function to automatically apply a module to extract the verbal emotions, which is developed to classify words into the six emotions (anger, disgust, fear, happiness, sadness, and surprise) based on the Facial Action Coding System (FACS).[5] The results of the verbal emotion analysis are displayed on our publicly available website as a graph, according to the temporal transition of deliberations, and a radar chart of the emotions of each speech (Figure 7).

To examine the correspondence between verbal and non-verbal emotions, we examine the parliamentary deliberations at the recent plenary sessions of both houses and the Budget Committee meetings, which tend to be more heated due to the interactive questioning and answering. Specifically, we cover the plenary session of the House of Representatives (10 meetings), the plenary session of the House of Councillors (10 meetings), the Budget Committee of the House of Representatives (10 meetings), and the Budget Committee of the House of Councillors (10 meetings) during the 213rd ordinary Diet session from January 26 to June 23, 2024 (150 days). Using the emotion extraction API described above, we calculated the frequency of words considered to be related to each emotion in each speech, indexed it as a relative percentage of the speech, and defined the emotion index for each speech as the emotion that exceeds the neutral point of 0.5. Table 1 summarizes the basic statistics for the 40 meetings analyzed and those by meetings.

---

[3] Hall-Lew (2010).
[4] Masuyama and Matsuda (2023).
[5] sentiment_ja2. Go et al. (2009).
https://github.com/sugiyamath/sentiment_ja2
https://www.tensorflow.org/datasets/catalog/sentiment140

Figure 6: Word Cloud Analysis

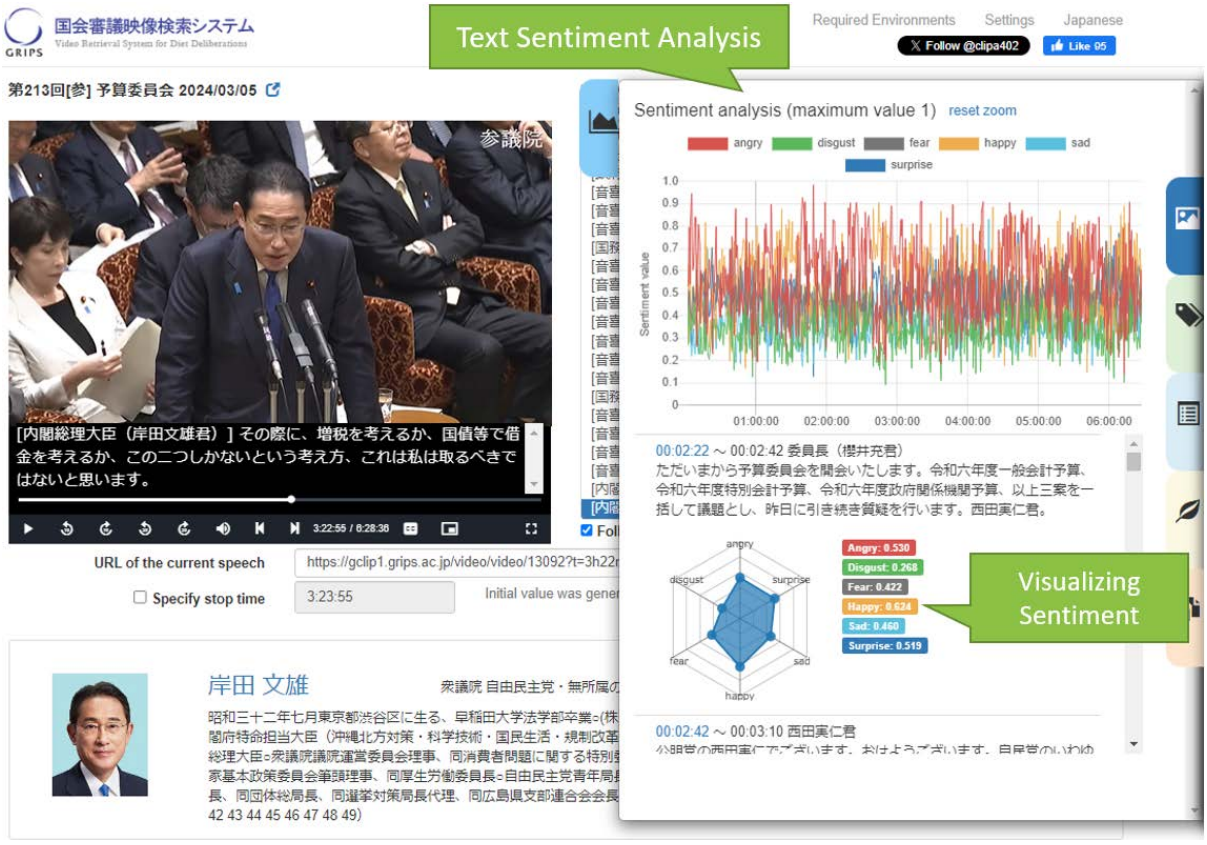


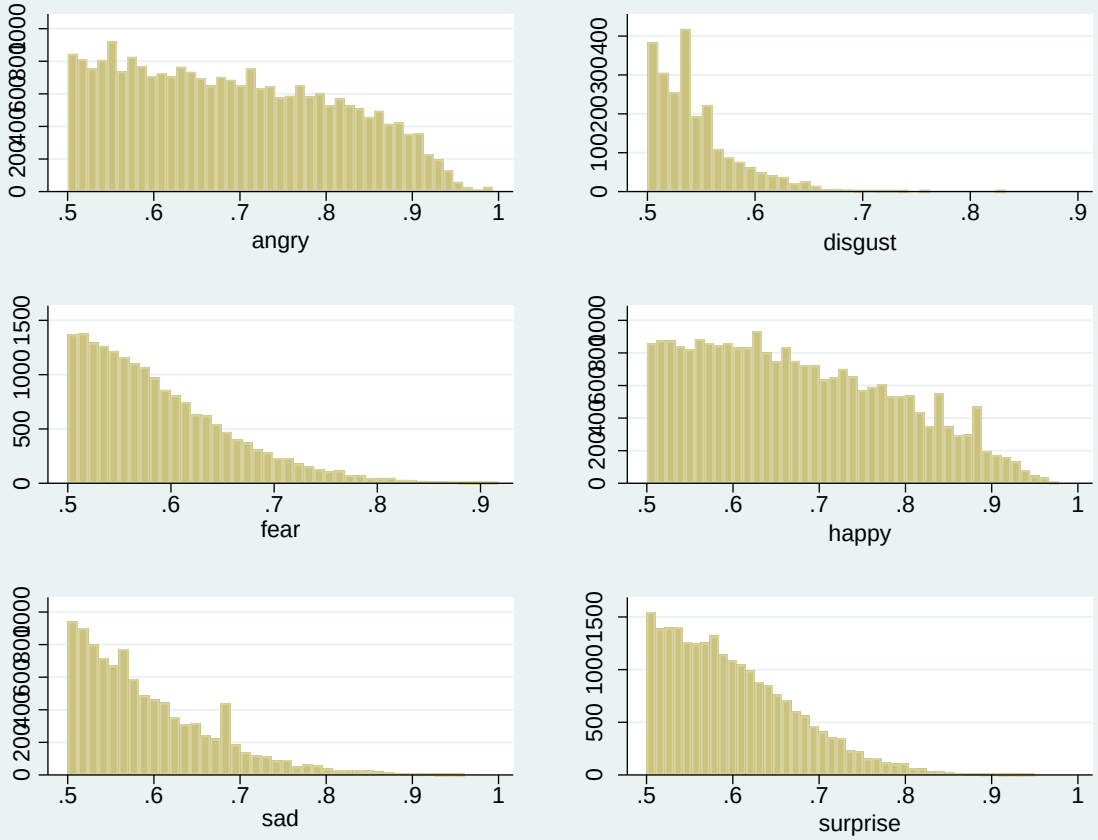
Figure 7: Text Sentiment Analysis

The total number of speeches in the target meetings was 43,964, with the Budget Committee approximately three times as many as the plenary session in both houses. The average number of characters for each speech was less than 60, indicating that the speeches made in the plenary session of the House of Representatives were somewhat shorter. The average score for anger is just below 0.7. Respectively, it is mid 0.5 for disgust, below 0.6 for fear, high 0.6 for happiness, and around 0.6 for sadness and surprise. Figure 8 shows the distribution of speeches according to the level of each emotion, indicating that there are many speeches at a relatively high level for anger and happiness, with some speeches in the 0.6 range for fear, sadness, and surprise, but only in the 0.5 range for disgust.

Table 1: Verbal Emotions in Speeches

		Obs	Mean	Std. dev.	Min	Max
#characters	All	43,964	58.1	42.3	2	517
	HR/Plenary	5,802	52.8	35.2	3	291
	HR/Budget	16,530	59.5	43.9	2	517
	HC/Plenary	6,062	59.3	37.7	4	403
	HC/Budget	15,570	58.2	44.4	2	514
angry	All	23,926	0.690	0.121	0.500	0.994
	HR/Plenary	3,752	0.709	0.122	0.500	0.994
	HR/Budget	8,232	0.685	0.120	0.500	0.993
	HC/Plenary	3,787	0.693	0.120	0.500	0.992
	HC/Budget	8,155	0.684	0.121	0.500	0.992
disgust	All	2,346	0.545	0.039	0.500	0.833
	HR/Plenary	323	0.549	0.042	0.500	0.738
	HR/Budget	819	0.544	0.040	0.500	0.833
	HC/Plenary	352	0.546	0.038	0.500	0.703
	HC/Budget	852	0.544	0.038	0.500	0.756
fear	All	18,499	0.590	0.071	0.500	0.917
	HR/Plenary	2,345	0.584	0.067	0.500	0.885
	HR/Budget	6,918	0.591	0.071	0.500	0.916
	HC/Plenary	2,386	0.587	0.073	0.500	0.905
	HC/Budget	6,850	0.591	0.071	0.500	0.917
happy	All	24,533	0.675	0.113	0.500	0.978
	HR/Plenary	2,936	0.658	0.106	0.500	0.961
	HR/Budget	9,565	0.680	0.116	0.500	0.978
	HC/Plenary	3,397	0.667	0.106	0.500	0.969
	HC/Budget	8,635	0.677	0.115	0.500	0.965
sad	All	9,827	0.594	0.078	0.500	0.961
	HR/Plenary	961	0.592	0.082	0.500	0.943
	HR/Budget	4,155	0.594	0.077	0.500	0.925
	HC/Plenary	870	0.587	0.082	0.500	0.961
	HC/Budget	3,841	0.597	0.078	0.500	0.939
surprise	All	22,425	0.601	0.074	0.500	0.950
	HR/Plenary	2,814	0.594	0.071	0.500	0.950
	HR/Budget	8,648	0.606	0.076	0.500	0.943
	HC/Plenary	2,879	0.590	0.067	0.500	0.932
	HC/Budget	8,084	0.603	0.075	0.500	0.942

Figure 8: Verbal Emotions by Speech



Note: Number of Speeches by Emotional Level

In our VRS, we are developing a facial expression analysis function based on facial recognition to grasp non-verbal emotions. The module developed for facial expression analysis makes it possible to extract a snapshot every 30 seconds from the deliberation video and convert the six emotions based on FACS by analyzing the speaker's facial image.[6] Therefore, we need some temporal approximation method to convert verbal emotion data comparable to the non-verbal emotion data extracted every 30 seconds. We can extract facial expressions at a point in time, while verbal emotions are the ratios in a time interval. Thus, as a first step to comparing verbal and non-verbal emotions, we converted the emotion indices for each speech to linearly interpolated indices every 10 seconds from the beginning of the meeting and matched up the non-verbal emotion data from the facial expressions within these 10-second intervals. Table 2 summarizes the emotion indices extracted from facial expressions every 30 seconds in the target meeting, normalized to the maximum value of 1 for each meeting. Table 3 shows the data obtained by linearly interpolating the emotion indices for each speech every 10 seconds, normalizing them with the maximum value of 1 for each meeting, and matching them with the emotion indices of the facial expressions in Table 2.

[6] DeepFace. Serengil and Ozpinar (2021).
https://github.com/serengil/deepface

Table 2: Non-verbal Emotions in Facial Expressions

		#Obs	Mean	Std. dev.	Min	Max
angry	All	21,990	0.127	0.207	0	1
	HR/Plenary	2,771	0.145	0.213	0	1
	HR/Budget	5,050	0.132	0.201	0	1
	HC/Plenary	6,622	0.116	0.212	0	1
	HC/Budget	7,547	0.127	0.205	0	1
disgust	All	21,990	0.014	0.086	0	1
	HR/Plenary	2,771	0.019	0.101	0	1
	HR/Budget	5,050	0.010	0.073	0	1
	HC/Plenary	6,622	0.024	0.110	0	1
	HC/Budget	7,547	0.005	0.058	0	1
fear	All	21,990	0.126	0.206	0	1
	HR/Plenary	2,771	0.100	0.167	0	1
	HR/Budget	5,050	0.097	0.176	0	1
	HC/Plenary	6,622	0.137	0.219	0	1
	HC/Budget	7,547	0.145	0.223	0	1
happy	All	21,990	0.032	0.136	0	1
	HR/Plenary	2,771	0.028	0.127	0	1
	HR/Budget	5,050	0.018	0.111	0	1
	HC/Plenary	6,622	0.038	0.132	0	1
	HC/Budget	7,547	0.038	0.157	0	1
sad	All	21,990	0.350	0.306	0	1
	HR/Plenary	2,771	0.402	0.294	0	1
	HR/Budget	5,050	0.291	0.267	0	1
	HC/Plenary	6,622	0.340	0.313	0	1
	HC/Budget	7,547	0.379	0.323	0	1
surprise	All	21,990	0.015	0.087	0	1
	HR/Plenary	2,771	0.020	0.100	0	1
	HR/Budget	5,050	0.005	0.056	0	1
	HC/Plenary	6,622	0.024	0.104	0	1
	HC/Budget	7,547	0.012	0.082	0	1

Table 3: Verbal Emotions (Interpolated) in Speeches

		#Obs	Mean	Std. dev.	Min	Max
angry	All	21,990	0.416	0.326	0	1
	HR/Plenary	2,771	0.505	0.320	0	1
	HR/Budget	5,050	0.355	0.323	0	1
	HC/Plenary	6,622	0.457	0.317	0	1
	HC/Budget	7,547	0.388	0.327	0	1
disgust	All	21,990	0.040	0.152	0	1
	HR/Plenary	2,771	0.047	0.165	0	1
	HR/Budget	5,050	0.033	0.135	0	1
	HC/Plenary	6,622	0.043	0.159	0	1
	HC/Budget	7,547	0.039	0.151	0	1
fear	All	21,990	0.299	0.303	0	1
	HR/Plenary	2,771	0.301	0.312	0	1
	HR/Budget	5,050	0.298	0.294	0	1

	HC/Plenary	6,622	0.295	0.307	0	1
	HC/Budget	7,547	0.304	0.301	0	1
happy	All	21,990	0.399	0.318	0	1
	HR/Plenary	2,771	0.344	0.319	0	1
	HR/Budget	5,050	0.417	0.312	0	1
	HC/Plenary	6,622	0.415	0.322	0	1
	HC/Budget	7,547	0.393	0.314	0	1
sad	All	21,990	0.141	0.241	0	1
	HR/Plenary	2,771	0.116	0.222	0	1
	HR/Budget	5,050	0.161	0.248	0	1
	HC/Plenary	6,622	0.115	0.226	0	1
	HC/Budget	7,547	0.158	0.254	0	1
surprise	All	21,990	0.363	0.314	0	1
	HR/Plenary	2,771	0.353	0.325	0	1
	HR/Budget	5,050	0.380	0.311	0	1
	HC/Plenary	6,622	0.358	0.316	0	1
	HC/Budget	7,547	0.359	0.311	0	1

Table 4: Correlation of Verbal and Non-verbal Emotions

	All	HR Plenary	HR Budget	HC Plenary	HC Budget
angry	0.013	-0.024	0.029	0.026	0.007
	(0.054)	(0.211)	(0.043)	(0.036)	(0.570)
disgust	-0.022	-0.032	-0.017	-0.029	-0.016
	(0.001)	(0.095)	(0.216)	(0.018)	(0.156)
fear	-0.004	0.030	0.008	-0.030	0.003
	(0.570)	(0.118)	(0.577)	(0.014)	(0.815)
happy	-0.005	0.015	0.074	-0.048	-0.014
	(0.468)	(0.427)	(0.000)	(0.000)	(0.217)
sad	-0.026	0.031	0.015	-0.081	-0.018
	(0.000)	(0.099)	(0.296)	(0.000)	(0.119)
surprise	-0.009	0.007	0.010	-0.028	0.001
	(0.164)	(0.727)	(0.479)	(0.022)	(0.950)

Note: Correlation coefficients and significance in parentheses.

Regarding non-verbal emotion indices based on facial expressions, on average, anger was higher in the House of Representatives than in the House of Councillors, while fear and happiness were higher in the House of Councillors. Disgust and surprise were higher in the plenary session than in the Budget Committee, while sadness tended to be higher in the plenary session of the House of Representatives and the Budget Committee of the House of Councillors (Table 2). The verbal emotion indices matched with the facial expression data were higher for anger and disgust in the plenary session and higher for sadness in the Budget Committee. Fear and surprise do not differ significantly, while happiness tends to be higher in the Budget Committee of the House of Representatives and the plenary session of the House of Councillors (Table 3). Thus, it is difficult to say that there is a consistent relationship between verbal and non-verbal emotions by meeting in the aggregate analysis. Although we must remember that there are limitations in the statistical analysis of such data, Table 4 summarizes the overall correlation between verbal and non-verbal emotions, indicating, at best, there is a weak positive relationship

for happiness in the House of Representatives Budget Committee. Although we cannot deny the possibility of a positive relationship for anger as well, it is difficult to say that there is at least a clear positive relationship between verbal and nonverbal emotions.

Conclusion

This paper has outlined our VRS, an attempt to utilize audio and visual information from parliamentary deliberations and legislative processes, departing from the tradition of focusing primarily on written records. Our VRS synchronizes the time information of deliberation videos with the text information of the meeting minutes and enables us to efficiently check the audio and visual information corresponding to each speech made in parliamentary deliberations. In this paper, we have explored how our VRS can generate data to compare verbal emotions in speeches and non-verbal emotions in facial expressions.

Our VRS aims to pinpoint the video segment corresponding to a speech from the parliamentary minutes, enable users to gain a visual understanding of the deliberation flow, and check the facial expressions of the speaker, which are unavailable from the written records. In addition, by adding subtitles to the deliberation video, it will be possible to use the deliberation videos even for those with hearing impairments, and by expressing the moment of speech as a URL, it will be easy to share the moment of interest in the deliberation video on the Internet through SNS. Moreover, our video retrieval system uses pattern recognition to extract images of supplementary materials and allows users to instantly check the graphical information, which is often a concise summary of the issues discussed in committee meetings.

Our VRS has great potential to boost the usage of parliamentary videos. The speech recognition techniques for creating timestamp data for matching video and text information apply to a wide range of meetings, including those of local assemblies, administrative councils, and international conferences, as well as other types of videos, such as TV news clips. Taking advantage of our VRS, which expresses the moment of speech as a URL and partially plays the video segments of interest, it may facilitate an experimental analysis of how visual information affects the understanding of parliamentary deliberations.

There is no doubt that the minutes are an essential source of parliamentary information, but the minutes do not record everything that happens in parliaments. Although the minutes are a valuable source of information as an official record, we must recognize that they are textual information that has somehow gone through the editing process, eliminating various non-textual information. For example, in everyday conversation, empathy and agreement are confirmed by eye contact and nodding. If the content conveyed in actual communication differs greatly depending on eye contact, gestures, vocal sounds, and voice, it is unsurprising that understanding of parliamentary deliberations differs whether reading the minutes or watching the videos. However, while parliaments are organizations that institutionally and systematically accumulate text, audio, and video information, there is a persistent bias in parliamentary research toward written records and audio and visual information has not yet been used systematically.[7]

[7] In parliamentary research, for example, Proksch and Slapin (2015) and Bäck and Debus (2016) have systematically analyzed parliamentary debates and speeches. A growing number of studies have shifted their focus from text to audio and video, such as Proksch et al. (2019), Dietrich et al. (2019), Proksch et al. (2019), Rheault et al. (2016), Rheault and Borwein (2019), Werlen et al. (2018), Werlen et al. (2021), and Schonhardt-Bailey (2022). However, such a

As an example of using our VRS, which enables us to check audio and visual information corresponding to speeches efficiently, we illustrated how our system can extract verbal and non-verbal emotions and suggested the possibility that the two are not necessarily in a positive relationship. Needless to say, we cannot deny that this possibility is due to the data limitations. Verbal emotion is indexed as the frequency of words associated with emotion in sentences and assumes a time interval as a unit of analysis, while facial expressions are based on snapshot images and indexed as a numerical value at a single point in time. The problem of the unit of analysis, whether a time segment or a single point in time, is fundamental. We use a linear interpolation of the frequency of words associated with emotion in a speech every 10-second interval and select the 10-second intervals that approximate the 30-second intervals at which facial expressions are extracted. It is undeniable that interpolation and time intervals are arbitrary. If we want to obtain a better temporal match, we may consider, for example, selecting words with solid emotional expressions, using them for keyword search, and extracting facial expressions from the snapshot when the speaker utters the word in question. Alternatively, it may be more reasonable to use the higher aggregation level as the unit of analysis to examine the differences in the speaker's characteristics, such as gender and party affiliation, or speech characteristics, such as whether it is a question or an answer.

Despite the limited analysis, the fact that we found a slight positive relationship for happiness in the House of Representatives Budget Committee may also suggest that it may be due to technical factors that reflect the difference in camera shooting techniques in the two houses: the camera focuses on the supplementary materials in the House of Representatives while keep shooting the speakers when they use the supplementary materials in the House of Councillors, making it challenging to extract facial expressions from smaller faces (Figure 5). In any case, the messages legislators try to convey through parliamentary deliberations may include not only what the words mean. Facial expressions and body language may be the opposite of words' meaning. There is a limit to understanding the diverse and multidimensional space of parliamentary deliberations through analysis of the minutes alone.

References

Bäck, H. and Debus, M. (2016) *Political Parties, Parliaments and Legislative Speechmaking*. Palgrave Macmillan.

Dietrich, B.J., Hayes, M., and O'Brien, D.Z. (2019) Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech, *American Political Science Review.* 113(4) 941-962.

Dumitrescu, D. (2016) Nonverbal Communication in Politics: A Review of Research Developments, 2005-2015, *American Behavioral Science*. 60: 1656-1675.

Go A., Bhayani, R., and Huang, L. (2009) Twitter Sentiment Classification using Distant Supervision, *CS224N Project Report*, Stanford, 1-12.

Hall-Lew, L., Coppock, E., and Starr, R.L. (2010) Indexing Political Persuasion: Variation in the Iraq Vowels, *American Speech*. 85: 91-102.

research area remains significantly less explored, and as Dumitrescu (2016) reviewed, the interaction between verbal and non-verbal information is understudied.

Kawahara, T. (2012) Transcription System Using Automatic Speech Recognition for the Japanese Parliament (Diet), *Proc. AAAI/IAAI,* pp.2224-2228.

Kawahara, T. (2024) Quantitative Analysis of Editing in Transcription Process in Japanese and European Parliaments and its Diachronic Changes, *ParlaCLARIN IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora*, Lingotto Conference Centre, Torino, Italy, May 20, 2024.

Masuyama, M. and Kawahara, T. (2019) Automatic Speech Recognition and Video Retrieval System for the Japanese Diet, *GRIPS Discussion Papers*. 19-09.

Masuyama, M., Kawahara, T., and Matsuda, K. (2024) Video Retrieval System Using Automatic Speech Recognition for the Japanese Diet, *ParlaCLARIN IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora*, Lingotto Conference Centre, Torino, Italy, May 20, 2024.

Masuyama, M. and Matsuda, K. (2023) Heteronyms in Diet Deliberations, (in Japanese) *GRIPS Discussion Papers*. 23-4.

Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016) Measuring Emotion in Parliamentary Debates with Automated Textual Analysis, *PLOS ONE.* 11(12) 1-18.

Rheault L. and Borwein, S. (2019) Multimodal Techniques for the Study of Affect in Political Videos, *PolMeth Conference*, MIT, Cambridge, MA, July 18-20, 2019.

Serengil, S. I. and Ozpinar, A. (2021) HyperExtended LightFace: A Facial Attribute Analysis Framework, *International Conference on Engineering and Emerging Technologies (ICEET)*, Istanbul, Turkey, 2021, pp. 1-4

Werlen, E., Moser, I., Imhof, C., and Bergamin, P. (2018) Is Reading Mirrored in the Face? A Comparison of Linguistic Parameters and Emotional Facial Expressions, *CEUR-WS.org*, 1-2226, paper2.

Werlen, E., Imhof, C., and Bergamin, P. (2021) Emotions in the Parliament: Lexical emotion analysis of parliamentarian speech transcriptions, *CEUR-WS.org*, 1-2957, paper10.

Proksch, S.O. and Slapin, J.B. (2015) *The Politics of Parliamentary Debate: Parties, Rebels and Representation*. Cambridge University Press.