Extending Search System based on Interactive Visualization for Speech Corpora

Tomoko Ohsuga, Yuichi Ishimoto, Tomoko Kajiyama
Shunsuke Kozawa, Kiyotaka Uchimoto, Shuichi Itahashi
National Institute of Informatics, National Institute for Japanese Language and Linguistics, The Open University
Gunosy Inc., National Institute of Information and Communications Technology, Tsukuba University
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
10-2 Midori-cho, Tachikawa City, Tokyo 190-8561, Japan
Walton Hall, Milton Keynes MK7 6AA, United Kingdom
Roppongi Hills Mori Tower, 6-10-1 Roppongi, Minato-ku, Tokyo 106-6125, Japan
4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan
osuga@nii.ac.jp, yishi@ninjal.ac.jp, tomoko.kajiyama@open.ac.uk
shunsuke.kozawa@gmail.com, uchimoto@nict.go.jp, itabashi.shuichi.da@alumni.tsukuba.ac.jp

Abstract

This paper describes a search system that we have developed specifically for speech corpus retrieval. It is difficult for speech corpus users to compare and select suitable corpora from the large number of various language resources in the world. It would be more convenient for users if each data center used a common specification system for describing its corpora. With the "Concentric Ring View (CRV) System" we proposed, users can search for speech corpora interactively and visually by utilizing the attributes peculiar to speech corpora. We have already proposed a set of specification attributes and items as the first step towards standardization, and we have added these attributes and items to the large-scale metadata database "SHACHI", then we connected SHACHI to the CRV system and implemented it as a combined speech corpus search system.

Keywords: speech corpus, retrieval, visualization

1. Introduction

Speech corpora are indispensable to speech research; several data centers of language resources have been set up worldwide to meet this demand by serving as a repository for language resources that include various speech corpora. They include the European Language Resources Association (ELRA), the Linguistic Data Consortium (LDC) in the U.S.A., the Chinese LDC / Chinese Corpus Consortium (CCC), the Speech Information Technology & Industry Promotion Center (SiTEC) in Korea, and the Speech Resources Consortium at the National Institute of Informatics (NII-SRC) / Language Resources Association (Gengo Shigen Kyokai; GSK) / Advanced Language Information Forum (ALAGIN) in Japan. The amount of data distributed from such data centers is very large. The diversity of the data gives users more freedom of choice, but it has become difficult to select suitable corpora for the intended purpose from the wide variety of corpora that have been made available. Although it is possible to search for these data from the website of each data center, the metadata of the corpus descriptions are not unified and also we cannot specify speech-specific conditions such as the recording environment in the search, so it is not easy for users to find the necessary corpus. Therefore, it would be more convenient for corpus users if the catalogue specifications of the corpora were standardized among all the various data centers and if the speech corpora could be retrieved by utilizing speech-specific conditions.

We have already proposed an interactive visualization and search system for speech corpora called the "Concentric Ring View (CRV)" system, which simultaneously creates a visual display while performing data retrieval (Itahashi, 2011). Using only a mouse, users can choose appropriate search keys for each of the attributes, and they can easily filter information by adjusting the keys. Retrieved results are displayed inside the rings, and users can filter and browse them in real time. On the other hand, a large-scale database system called "SHACHI" [1] was developed to collect metadata, such as tag sets, formats, and recorded contents, from language resources worldwide (Tohyama, 2008). To combine these two systems, we have revised corpus specifications of SHACHI and added attributes peculiar to speech corpora (Itahashi, 2014). This paper reports the revised version of the CRV system obtained by incorporating about 1200 items of speech corpus information from the SHACHI system.

In the following, we give an outline of the CRV system in Section 2, followed by an outline of the SHACHI metadata database and the attributes of the speech corpus specifications in Section 3. Section 4 presents the revisions of the CRV system, Section 5 describes an experiment carried out to verify the effectiveness of the revised system and its results, and finally, Section 6 concludes the paper.

2. Concentric Ring View (CRV) System

One of the authors previously developed a novel search and display system called the CRV system. This system is an interactive environment for integrating searching and browsing of various items, and its effectiveness has already been shown in applications such as a search system for iPhone applications, a pictorial book of flora, and so forth (Kajiyama, 2014). The CRV system is composed of several concentric rings, each of which corresponds to a selected attribute. The authors have adopted this system as the search system for speech corpora.

[1] http://shachi.org/

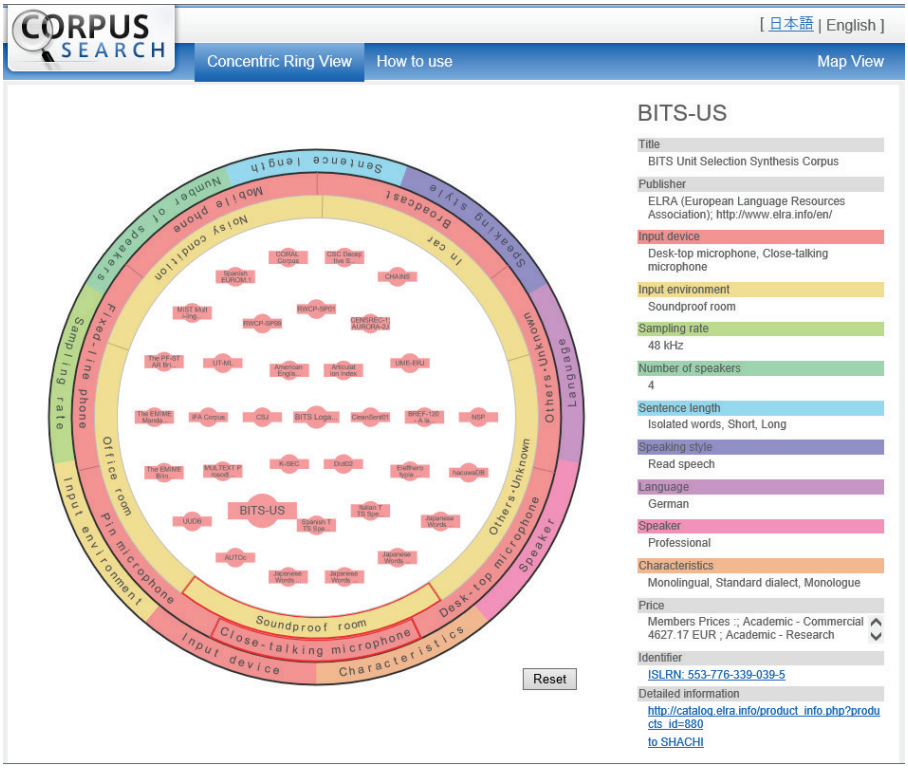


Figure 1: Screenshot of CRV displaying search results for corpora specifying “Close-talking microphone” for “Input device” attribute and “Soundproof room” for “Input environment” attribute, and selecting the “BITS-US” corpus from the corpora in the rings to display its detailed information.

Figure 1 illustrates an example of a screenshot during a search for speech corpora. Initially, only the outermost ring is displayed, which expresses the attributes of the corpora. It is divided into several sectors, each corresponding to an attribute. By clicking a certain sector, another ring, an item ring, appears inside. This item ring contains the category items that correspond to the attribute category specified on the attribute ring. The item ring shown inside has the same color as that of the corresponding sector of the outermost attribute ring. These corpora specified by the attribute and item on the rings are displayed inside the rings. A user can rotate the item ring and adjust the item by dragging a suitable sector of the item ring and browse the search results shown inside the rings. The current item is always shown at the bottom of each ring in a highlighted sector, so the user can easily check the current position or condition. The displayed information can be narrowed down by specifying more attributes, which causes other rings to appear inside. This results in AND retrieval of the specified items, and it is also possible to perform AND retrieval by displaying multiple rings of the same attribute. The search results do not depend on the order of rings displayed. By clicking a displayed ring again, the ring disappears and its retrieval condition can be canceled. This technique allows users to easily and precisely specify each item. The details of a specified corpus are displayed on the right of the screen by clicking the desired corpus shown inside the rings. This is an attribute-based search and users can search for corpora by any attribute in any order.

Table 1 shows the list of attributes that can be used as the retrieval key and the items of each attribute. Initially, only the attribute ring on the left of the table is displayed; by clicking the necessary attribute sector, the corresponding item on the right of the table will be displayed inside the ring. The attribute of “Language” indicates the geographical area of the languages, such as Asia, Europe, and Africa, which were originally used in SHACHI. The specific language name is shown in the detailed information area on the right of the screen, as shown in Fig. 1. Further, specific values for the “Sampling rate” or “Number of speakers” are displayed in the “Detailed information” area.

Of course, many data centers have their own corpus retrieval system that can search the corpora they provide, but they cannot search the corpora by specifying the recording environment or speaking style as shown in Table 1. For example, the items that can be used in the LDC catalogue include “Publication name”, “Author”, “Catalog number”, “Language(s)”, “Member year(s)”, “DCMI type(s)”, “Data source(s)”, “Research project(s)”, “Recommended application(s)” and free keywords. META-SHARE[2], which is an open network of repositories for sharing and exchanging language data and tools, can search by selecting “Media type”, “Language”, and other criteria. CLARIN - European Research Infrastructure for Language Resources and Technology[3] provides a search system using facets including “Resource type”, “Modality”, and “Format”. In contrast, the CRV retrieval system proposed in this paper can retrieve speech corpora utilizing

[2] http://www.meta-share.org/

[3] https://www.clarin.eu/

Table 1: Set of attributes and items for retrieval of speech corpora

Attribute	Item
Input device	Desk-top microphone, Close-talking microphone, Lapel microphone, Fixed-line phone, Mobile phone, Broadcast, Others/Unknown
Input environment	Soundproof room, Office room, Noisy condition, In-car, Others/Unknown
Sampling rate	SR < 10 kHz, SR < 20 kHz, 20 kHz ≦ SR, Unknown
Number of speakers	No < 10, No < 100, No < 1000, 1000 ≦ No, Unknown
Sentence length	Isolated words, Short, Long, Others/Unknown
Speaking style	Read speech, Acted speech, Spontaneous speech, Others/Unknown
Language	Japan, Asia, Europe, Africa, America, Oceania, Others/Unknown
Speaker	Non-native, Professional, Child, Senior, Others/Unknown
Characteristics	Multilingual, Dialect, Dialogue, Emotional, Non-speech, Others/Unknown

speech-specific retrieval conditions; another advantage is that users can manipulate the system visually and that they can add or modify retrieval conditions interactively. In particular, it is useful when users do not know the corpus name and when they want to search corpora suitable for their own objectives or experimental conditions.

3. SHACHI, Large-scale Metadata Database of Language Resources

SHACHI is a large-scale metadata database of language resources developed jointly by the National Institute of Information and Communications Technology (NICT) and Nagoya University of Japan (Tohyama, 2008). It collects detailed metadata information on language resources worldwide, including the resources provided by ELRA, LDC, CCC, SiTEC, NII-SRC, and so forth.

The metadata set adopted by SHACHI conforms to the OLAC[4] metadata set, which is based on 15 fundamental elements of the Dublin Core[5] and constitutes an extended version of OLAC with originally added metadata elements that were considered to be indispensable for describing the characteristics of language resources. It would be more convenient if we could use a set of attributes that describes a corpus to search through the various speech corpora. However, SHACHI mainly focuses on text corpora and does not have sufficient suitable attributes to describe speech corpora. Therefore, we have revised the set of attributes and items. In detail, we have added attributes such as "Input device", "Input environment", "Sampling rate", "Number of speakers", and "Speaking style" to SHACHI (Itahashi, 2014). The attributes of "Sentence length", "Language", and each item of "Speaker" and "Characteristics" were originally used in SHACHI. At present, there are 55 metadata elements in the SHACHI system.

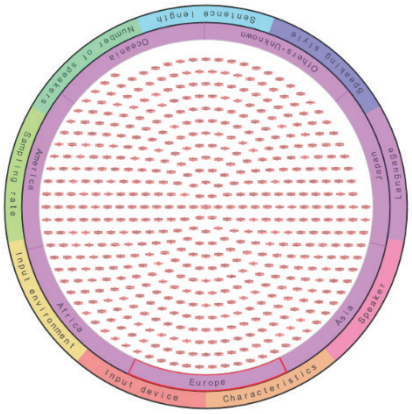


Figure 2: Screenshot of CRV displaying search results for corpora specifying "Europe" for the "Language" attribute.

SHACHI contains 3300 compiled language resources as of Sep. 2017, including 1214 speech corpora. ELRA also has a similar system, i.e., a universal catalogue[6]. It also has its own search function based on keywords, for instance, one can find 749 products among 1645 corpora by specifying the "Speech" keyword as of Sep. 2017. The most important feature of SHACHI is that the automatically collected metadata are manually corrected and thus error-free (Tohyama, 2008). We think that SHACHI is the most suitable database for integration with the search system as it has been extended by adding attributes characterizing speech corpora such as the speech recording environment and speaking style.

4. Connecting SHACHI with CRV Search System and Revision of the Search System

Because it would be much more convenient if SHACHI could be combined with a visual search system such as CRV, we tentatively developed software for extracting the corpus metadata from SHACHI and transferring them to the CRV system.

We confirmed that it is possible to retrieve target corpora using the proposed system in a preliminary estimation experiment using 50 corpora (Itahashi, 2011). We have now incorporated all 1214 speech corpora enrolled in SHACHI in the CRV system. After increasing the number of corpora to be retrieved, however, it turned out that the existing system had a problem in the user interface. The size of the retrieved corpus icons to be displayed inside the ring was fixed in the previous system, and so when there

[4] http://www.language-archives.org/

[5] http://dublincore.org/

[6] http://universal.elra.info/

were many corpora to be displayed, they were divided into multiple pages and the next page could be displayed by rotating the ring. However, users may have found it difficult to understand that the operations of changing the retrieval keys and showing the next page are performed by the same ring operation in the case that each attribute has discrete values. Also, it was not easy to grasp the number of retrieved corpora. Thus, we have devised a new system that can display all the retrieved corpora in a single page by reducing the size of each corpus icon. When there are too many retrieved results, each corpus icon is illustrated with a very small size, suggesting that the user should add more retrieval conditions to reduce the number of retrieved corpora. Figure 2 shows an example of retrieved results when "Europe" is specified as the "Language" attribute. It can be seen that there are too many retrieved results and that it is necessary to add more attributes to reduce the number of retrieved results.

This modification makes it possible for users to intuitively grasp the approximate number of retrieved corpora corresponding to the retrieval items by rotating the ring.

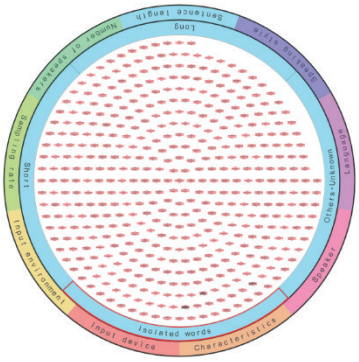
Moreover, we have also added identifiers such as the International Standard Language Resource Number (ISLRN) and the Digital Object Identifier (DOI) as links to each landing page. These items are displayed in the detailed information area, as shown in Fig. 1. Incidentally, detailed information such as the formal title, the publisher, the price, and the URL originally existed in SHACHI, but the ISLRN and DOI links have been newly added to the SHACHI system as one of the present modifications.

5. Verifying the Extended Search System

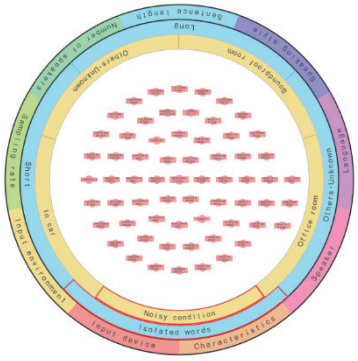
We performed a retrieval experiment to verify the effectiveness of the revised search system. Here we show the process of retrieving corpora with read-aloud isolated words under a noisy condition. First, the "Sentence length" attribute in the screenshot of Fig. 3(a) was clicked, and then the inside ring was rotated until the "Isolated words" item reached the bottom as shown in Fig. 3(b). Next, the "Noisy condition" item was selected as the "Input environment" attribute as shown in Fig. 3(c). Because quite a large number of corpora still remained, we added the item "20 kHz ≦ SR" for the "Sampling rate" attribute, which reduced the number of retrieved corpora to four, as shown in Fig. 3(d).

We conducted a preliminary questionnaire survey of 15 people (nine males and six females) who have experience of using speech corpora. After completing a simple retrieval task, we asked them if the proposed system was useful, for which obtained an average value of 3.8 on a five-step scale from one to five. This figure is better than the survey result of 3.2 for our former MDS-based retrieval system (Yamakawa, 2008). The result shows that this system is useful and effective for retrieving speech corpora. We also obtained an average score of 3.5 for "Relevance of attributes", with some comments that it is not easy to predict what items exist in each attribute. We plan to add a "Show attribute" button on the screen so that users can see the table of attributes any time they want. As future work, enrichment of the metadata is necessary to reduce the number of corpora classified into the "Unknown" category for each attribute. Also it is desirable to smoothly navigate among the attributes that should be chosen next for effective retrieval.

(a): Initial screenshot showing only attribute ring.



(b): Specifying "Isolated words" as "Sentence length".



(c): Specifying "Noisy condition" as "Input environment".

(d): Specifying "20 kHz ≦ SR" as "Sampling rate".

Figure 3: Screenshots of CRV display during search process.

6. Conclusions

We have proposed corpus specification attributes and items to give corpus users easy access to the speech corpora catalogues, and we connected SHACHI to our search system for speech corpora. As the search system, we adopted the Concentric Ring View (CRV) system, which simultaneously searches for and displays various objects. Users can search speech corpora interactively and visually by utilizing the attributes peculiar to speech corpora as a web application system. The present system can be accessed at the URL indicated at the end of this paragraph. We plan to continue adding more corpora and to carry out a more elaborate assessment of the proposed system by user questionnaire.

http://corpus-search.nii.ac.jp/ring/

7. Bibliographical References

Itahashi, S., Kajiyama, T., Yamakawa, K., Ishimoto, Y. and Matsui, T. (2011). Interactive visualization search system for speech corpora. *Proc. Oriental COCOSDA 2011*, pp. 157–161.

Itahashi, S., Ohsuga, T., Ishimoto, Y., Kojima, H., Uchimoto, K. and Kozawa, S. (2014). Revised catalogue specifications of speech corpora with user-friendly visualization and search system. *Proc. Oriental COCOSDA 2014*, pp. 60–64.

Kajiyama, T. and Satoh, S. (2014). An interaction model between human and system for intuitive graphical search interface. *International Journal of Knowledge and Information Systems*, 39(1) : 41–60.

Tohyama, H., Kozawa, S., Uchimoto, K., Matsubara, S. and Isahara, H. (2008). Construction of a metadata database for efficient development and use of language resources. *Proc. LREC 2008*, pp. 1687–1692.

Yamakawa, K., Matsui, T. and Itahashi, S. (2008). MDS-based visualization method for multiple speech corpora. *Proc. INTERSPEECH2008*, pp.1666–1669.