Paper:

Text-Style Conversion of Speech Transcript into Web Document for Lecture Archive

Masashi Ito*, Tomohiro Ohno, and Shigeki Matsubara*****

*Graduate School of Information Science, Nagoya University
**Graduate School of International Development, Nagoya University
***Information Technology Center, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
E-mail: {ohno, matubara}@nagoya-u.jp
[Received November 25, 2008; accepted March 25, 2009]

It is very significant to the knowledge society to accumulate spoken documents on the web. However, because of the high redundancy of spontaneous speech, the faithfully transcribed text is not readable on an Internet browser, and therefore not suitable as a web document. This paper proposes a technique for converting spoken documents into web documents for the purpose of building a speech archiving system. The technique edits automatically transcribed texts and improves their readability on the browser. The readable text can be generated by applying technology such as paraphrasing, segmentation, and structuring transcribed texts. Editing experiments using lecture data demonstrated the feasibility of the technique. A prototype system of spoken document archiving was implemented to confirm its effectiveness.

Keywords: natural languages, spoken language processing, digital archiving, web contents, paraphrasing

1. Introduction

The enormous amount of information on the web has been forming an infrastructure for the so-called knowledge society. Most of such information is expressed in text at the present time. Hereafter, it is expected that not only text but also audio, video and other mixed media will be utilized as knowledge resources. Above all, speech is a kind of language media, and produced daily. In fact, the amount of spoken documents being produced is overwhelmingly more than that of written documents. If speech such as discourses, lectures, monologues, conversations, and discussions were accumulated on the web, the amount of sharable information in the knowledge society would get much larger.

Several styles for archiving the spoken documents on the web can be considered. Though it seems that a style of uploading speech data would be very simple and would have high feasibility, its significance from a viewpoint of reusability is not so large. If the transcribed text is also archived in addition to the speech data, its effectiveness would be increased in terms of accessibility. On the other hand, because of the high redundancy of spontaneous speech, the transcribed text itself is not readable on an Internet browser, and therefore not suitable as a web document.

This paper proposes a technique for converting spoken documents into web documents for the purpose of building a speech archiving system. The technique edits automatically transcribed texts and improves their readability on the browser. The readable texts can be generated by applying language technology such as paraphrasing, segmentation and structuring to the transcribed texts. An edit experiment was conducted by using lecture speech data, and the result has shown the feasibility of the technique. Furthermore, a prototype system of spoken document archiving was implemented in order to confirm its effectiveness.

This paper is organized as follows: Section 2 discusses spoken document archiving on the web. Section 3 describes an edit technique of spoken documents. Section 4 reports evaluation of the technique. Section 5 details our speech archiving system. Section 6 presents conclusions.

2. Spoken Document Archiving on the Web

Our objective is to construct an environment in which spoken documents are archived as web documents. A web document is usually described using markup language such as HTML, and created for the readability on a browser. In that sense, a spoken document is not suitable for being read with an Internet browser. Therefore, a transcribed text needs to be converted into an HTML text.

Figure 1 shows a spoken document archiving system. An input speech is transcribed by automatic speech recognition and the text is converted into a text which is easy to read by sentence-style conversion and so on. Finally, a synthesized speech is generated.

In work done on speech archiving for Web content creation, Bain et al. have created a system named ViaScribe in order to improve accessibility of speech data [1]. ViaScribe targets lecture environments and creates learning material integrating audio, video, slides, and transcribed

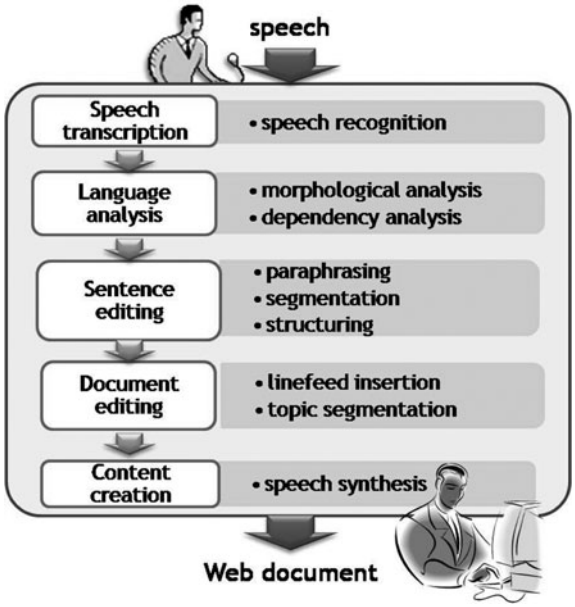

Fig. 1. Spoken document archiving system.

texts that are uploaded to the Web for access via a browser. ViaScribe inserts linefeeds where pauses occur in transcriptions to improve readability. Bain et al. targeted English, which is far less redundant than a language such as Japanese. Therefore in Japanese the readability could not be provided by only linefeed insertion based on pauses.

To improve the readability of a text based on the discourse structure, Shibata et al. have proposed a method of automatically generating summary slides from a text. They have clearly specified the relations between sentences or clauses by itemization. However, their research differ from ours in that the units such as words or phrases are not targets. Furthermore, as treating the written language, they do not consider the features of spoken language.

As a technique for converting transcribed text, there exists research on automatic summarization [3–6] and conversion from spoken language to written language [7]. However, they are all conversions between plain texts. In contrast, we propose a method for converting transcribed text into a web document, which can be utilized as a fundamental technology in the archiving system.

3. Text Editing of Spoken Document

The conversion of spoken documents into web documents is executed by editing spoken sentences so that they become easy to read. The conversion is realized by performing paraphrasing of expressions, segmentation into short sentences and structuring of a sentence in sequence. In addition, basic linguistic annotations are provided to each sentence by morphological analysis, clause boundary analysis and dependency analysis. In what follows, we detail each processing which constitutes the sentence-style conversion.

Table 1. Major types of paraphrased expressions and their frequency.

Polite expressions (e.g. しておりました (did) ⇒ していた)	112
Spoken language (e.g. おっきい (big) ⇒ おおきい)	51
Honorific verb (e.g. させて頂く (do) ⇒ する)	32
Collocation "*toyu*" (e.g. 正常化ということが行われた (perform the normalization)⇒ 正常化が行われた)	26
Completion and correction (e.g. 私普段から (I usually)⇒ 私は普段から)	27
Prefix (e.g. お料理を食べる (eat foods)⇒ 料理を食べる)	14

3.1. Paraphrasing

In spoken language, there are a lot of expressions which are not proper as written language expressions. Such expressions can be converted into readable expressions by paraphrasing them into other expressions.

In order to create the paraphrase rule for automatic sentence editing, we paraphrased the transcribed lecture speech text by hand. We used 100 sentences extracted from seven lectures in which the speaker was different. A total of 329 paraphrases were performed. **Table 1** shows the breakdown of the paraphrased expressions.

Paraphrasing involves deletion, replacement, or insertion, depending on the type of expression. In *deletion*, redundant expressions peculiar to spoken language are removed. There exists not only fillers (uh, er etc.) and hesitations but also other redundant expressions. In the following example,

(1) 正常化 ということ が行われた (performed the thing which is called normalization)

"ということ (the thing which is called)" is redundant and deleting it improves readability.

In *replacement*, expressions which are not proper as written language are replaced with other expressions. In the following example,

(2) 日本の本 なんか を読みました (I read Japanese books, etc.)

"なんか (etc.)" is an expression particular to spoken language. By replacing "なんか (etc.)" with "など (etc.)", it becomes a proper expression. Furthermore, in honorific expressions, polite expressions, and so on, there are several cases where replacement is needed.

In *insertion*, expressions which do not appear in spoken language are added.

In order to implement the three types of paraphrasing mentioned above, we created paraphrasing rules by using the actual lecture speech data [8]. Here, in insertion, since considering the meaning is required, it is difficult to create the rules by surface information alone. Therefore, the target of this paper is limited to deletion and replacement. **Fig. 2** shows examples of the created rules and their applications.