MEMORABLE SPOKEN QUOTE CORPORA OF TED PUBLIC SPEAKING

Fajri Koto[†‡], Sakriani Sakti[†], Graham Neubig[†], Tomoki Toda[†], Mirna Adriani[‡], Satoshi Nakamura[†]

[†]Nara Institute of Science and Technology, Japan
[‡]Faculty of Computer Science, Universitas Indonesia, Indonesia
fajri91@ui.ac.id, mirna@cs.ui.ac.id, {ssakti,neubig,tomoki,s-nakamura}@is.naist.jp

ABSTRACT

In this paper we present the construction and analysis of memorable spoken quote corpora from TED public speaking. Memorable quotes are interesting and useful words which usually contain generic pearls of wisdom that could achieve public awareness and retained in people consciousness. Our study aims to reveal why can some public speeches can be retained in people mind and make their consciousness to like and share it, while some others can not. To achieve this purpose, the relevance corpora is required to perform system quantitative evaluation. In this study, we start with the collection of the corpus from TED public speaking. Specifically, we utilize 899 video files of TED Talks and more than 2000 speech quotes annotated by TED team. We then complement the data with non-memorable quotes. According to shares number of quotes which are provided by TED, we also annotate memorable quotes with popularity factor. Analysis of memorable spoken quotes is done based on speech duration, F0, and popularity.

Index Terms— memorable quote, public speaking, spontaneous speech, corpora, favorable

1. INTRODUCTION

Research about speech processing has been actively investigated over decades. Specifically, study related to dialog system [1], speech recognition [2], speech summarization [3], and speech synthesis [4, 5] over languages are expanding along with relevance corpora which are successfully collected. One of the goal is to build spoken dialog system which enable machine to interact with human naturally. Consequently, understanding human conversational expressiveness to social presence that may gain partner acceptance becomes important. The term expressiveness used in this work does not specifically refer to emotional expressiveness, but to describe the skill of communicating genuine involvement in the conversation, including the choice of words and the way it phrased (i.e., loudness and intonation). Enhanced expressiveness may contribute to dramatic effect, making the message easier to listen to. Here, we focus on studying human expressiveness during public speeches, in which how the important messages are conveyed that may be retained in audience consciousness.

Memorable quotes are defined as interesting and useful words which usually contain generic pearls of wisdom expressed with unusual combination of words in ordinary sentences [6]. Through history, the best speeches of all time normally feature memorable quotes that genuinely inspire the audience. For instance, the most famous quote of John F. Kennedy: *"Ask not what your country can do, ask what you can do for your country"*. History has proven the existence of this memorable quote which inspired many generations since John F. Kennedy gave this speech in January 1961[1].

Nowadays, one popular site in of public speeches is TED[2]. TED started out in 1984 as a conference bringing together people from three worlds: Technology, Entertainment, Design. TED talks bring together the world's most fascinating thinkers and doers, who are challenged to give the talk of their lives in about 5-25 minutes. Many famous people have given speeches on TED and inspired people by their memorable words. Recently, TED has started "TED Quotes," which collects memorable quotes from TED talks, annotates them manually, groups them by category, and provides an easy way for people to share their favorite quotes. The most popular quotes can have more than a thousand shares.

We initiate our study by collecting corpus of memorable quote from TED public speaking. Specifically, in this study we collect more than 2000 spoken quotes and 899 corresponding TED Talks video that have been manually annotated. We build the segmented corresponding audio file and complement the corpus with non-quotes spoken data that are randomly generated. Manually checking of 899 subtitle/transcription files were also done. According to shares number of quotes which are provided by TED, we also annotate memorable quotes with favorableness factor. Analysis of memorable spoken quotes is done based on speech duration, F0, and popularity.

The rest of this paper is structured as follows. Section 2 summarizes some related works. Section 3 provides the procedure of data construction. The analysis of data regarding

[1]http://ushistory.com
[2]http://www.ted.com/

to duration, F0 and favorableness will be given in Section 4. Finally, conclusions are drawn in Section 5.

2. RELATED WORK

Study related to natural expressive speech has been done by some researchers. Bulut et al. work in synthesizing four emotional states: anger, happiness, sadness, and neutral using a concatenative speech synthesizer [4]. Eide et al. add five speaking styles - neutral declarative, conveying good news, conveying bad news, asking a question, and showing contrastive emphasis - in synthesizing speech [5]. Generating expressive speech for storytelling applications were also done by Theune et al. They designed and implemented a set of prosodic rules for converting neutral speech into storytelling speech [7]. However, most works related to synthesizing emotional expressiveness. Here, we study on memorable spoken quotes; the skill of communicating genuine involvement in the conversation, including the choice of words and the way it phrased (i.e., loudness and intonation) that may be retained in audience consciousness.

Research related to memorable quote is still very limited. There is only one study that has been published and discuss about memorable quote in text document. Bandersky et al., extracts some text features in order to analyze how a phrase in book can be memorable [8]. This research stems to the fact that there are close to 130 million unique book records in the world libraries today, an many of these are being digitized [9]. Moreover, many annotated text quotes are spread in Internet today. For instance, BrainyQuote[3], and WikiQuote[4] which have been developed to provide several inspirational groups of quotes from many resources.

Another study by Danescu-Niculescu-Mizil et al. [10] attempted to investigate the effect of phrasing on a quote's memorability from movie transcription. They argue that quotes differ not only in how they are worded, but also in who said them and under what circumstances. Although this study focused on spoken words, the work is limited to only textual data of movie transcription. While most techniques developed so far for memorable quote detection have focused primarily on the processing of text, we are interested in discovering memorable spoken quotes of real public speeches.

3. DATA CONSTRUCTION

As described in Section 1, the memorable spoken quotes corpora were built by utilizing TED speech, manually annotated quote, and transcription file of corresponding video speech. In total there are 2152 annotated quotes by July 2013 in TED website. They are required 914 speeches with its corresponding transcriptions to be processed. Due to there were not 15 needed transcription files, we reduced 34 quotes and the rest are 2118 memorable spoken quotes, with 899 required audio files.

At Fig. 1 we present the stage of corpora construction. First we downloaded all required file: 2118 memorable spoken quotes, 899 TED speeches and their transcriptions. We then manually checked all 899 transcription files and found there were some of them which had time mismatch for 1-10 seconds. The details of this mismatch of transcription are summarized in Table. 1.

Table 1. *The statistic of time mismatch for all transcription file.*

time mismatch	Count
-4 seconds	1
-3 seconds	202
0 second	660
2 seconds	1
3 seconds	31
6 seconds	2
10 seconds	2
Total	899

After manually checking and updating the transcriptions, we find the segment timing of every memorable spoken quotes in their transcription file. Non-quote data are then randomly generated to complement the corpora with explanation as follows: 1) The length of non-quote data were randomly generated in range 1-3 passages of transcription. We consider this case based on the length of existing quote data, 2) For each speech, we generate non-memorable quote as many as existing quotes in that speech. After we complement the data, segmentations then were applied and the data are ready to be extracted and analyzed.

4. ANALYSIS

4.1. Duration Analysis

In Table. 2 we provide the statistic of memorable quote corpora. We divide the speeches based on Speech-Duration Interval (SDI) in minute unit. It shows that 42.37% of our speeches corpora (386 speeches) have duration interval in 15-20 minutes, while 32.7% lay on 20- interval and the rests are in 0-15 interval. The quote utterances for each interval are also provided. We present Fig. 2 to show the changing of quote utterance in each interval. The quotes utterance normally increase from lower interval, then achieve maximum number of utterance in 15-20 interval. But, it then suddenly decreases when it reaches data with SDI greater than 20 minutes. In average, there will be two memorable segments in a speech interval which will be recognized as spoken quote by public consciousness.

[3]http://www.brainyquote.com
[4]http://www.wikiquote.com

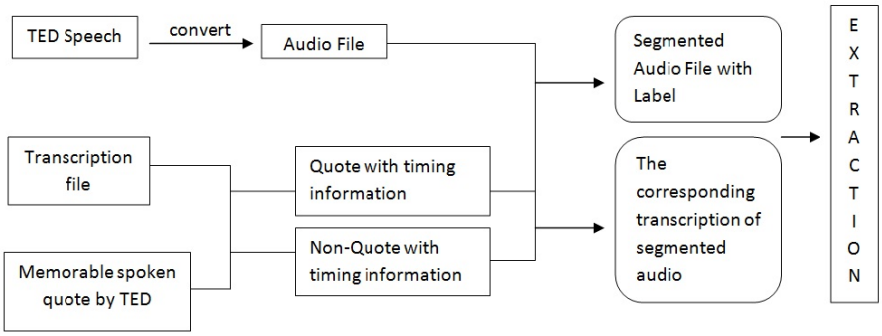

Fig. 1. The construction of memorable spoken quote corpora.

Table 2. *The statistic of quote utterance in speech corpora*

SDI (min)	# TED Talks	# Quotes in TED Talks	Quotes Avg Dur (sec)	Quotes Position		
				q1	q2	q3
0-5	71	123	11.028	36	40	47
5-10	147	295	11.224	80	88	127
10-15	187	443	10.673	141	139	163
15-20	386	963	11.308	366	288	309
20-	294	294	11.976	110	92	92
Total	899	2118	11.236	733	647	738

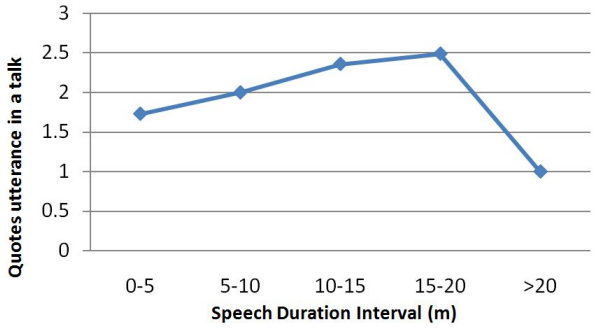

Fig. 2. The average number of Quotes in each TED Talk interval.

Quotes Avg Dur (sec) in Table. 2 represents the duration average of quote utterance for each interval in second unit. Our data reveal that quote utterances have similar duration for all SDI, about 10-11 seconds. Starting position of these spoken quote utterances are also our concern in this section. In Table. 2 we divide starting position of quote utterance in three segments. The first one-third segment of speech is denoted as q1, while q2 and q3 are the second and the third of next one-third segment. We then count the utterance of spoken quote for each segment and the results give ratio $q1 : q2 : q3 = 1.13 : 1 : 1.14$ for their total. It reveals that the utterances of memorable quote can be spoken in any duration of speech and can not be easily determined only based on their starting position.

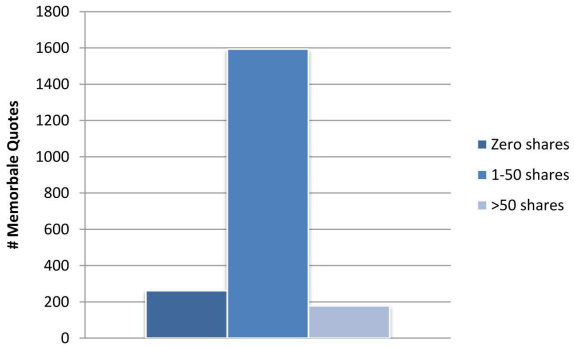

Fig. 3. Memorable quote data distribution according to shares number

4.2. Popularity analysis

TED provide an easy way for user to share their favorite quotes. The shares number of every quotes are publicly provided by TED. From the total of 2118 memorable quotes, the popular quotes can have more than a thousand shares, while the non-popular quotes have zero shares. For example, the most popular quote in our corpus is "*If you hire people just because they can do a job, they will work for your money. But if you hire people who believe what you believe, they will work for you with blood and sweat and tears*" given by Simon Sinek. This quote was shared by 4788 people. However, only very few quotes are shared by more than thousand people, while a large number of memorable quotes are shared by around 1-50 people (See the distribution in Fig. 3).

In this preliminary study, we only focused on extreme cases and constructed a corpus with memorable quotes that have zero shares (labeled as non-popular quotes), and memorable quotes that have more than 50 shares (labeled as popular quotes). Here, all new published quotes still have zero shares, and thus they are excluded from data as it may not be irrelevant to annotate them as non-popular quotes. In total, the corpus consists of 262 non-popular quotes and 179 popular quotes.

4.3. F0 Analysis

Danescu-Niculescu-Mizil et al. in their work argue that there may be factors which make information retained in people consciousness. One of those factors is may be due to the way of it is expressed. In emotional speech, F0 has also been investigated and stated as important feature [11]. The study by Liscombi et al. found that higher F0 may correlates with positive-action emotion[12].

Table 3. *F0 comparison between both corpora*

F0	Quote	Non-Quote
F0-Max	343.39	323.81
F0-Min	49.83	52.47
F0-Range	293.57	271.34
F0-Mean	169.61	168.42

In this preliminary study, we investigate F0 features between memorable and non-memorable spoken quotes. Table.3 presents F0 analysis between memorable and non-memorable quotes. Based on INTERSPEECH 2009 paralinguistic challenge configuration (IS09 Paraling features) [13], we extract F0-Max, F0-Min, F0-Range and F0-Mean. It is done using *openSMILE*[5]; a feature extraction toolkit, which unites feature extraction algorithms from the speech processing and the Music Information Retrieval communities [14].The result shows that F0-Mean of memorable quotes apparently are higher than non-memorable quotes. This may indicate that people tend to act in positive-action emotion in emphasizing important content during public speeches. Furthermore, as F0-Range (F0-Min:F0-Max) of memorable quotes is larger than non-memorable quotes, it may also reveal a tendency that memorable quotes are spoken with more variative intonation.

5. CONCLUSION AND FUTURE DIRECTION

In this paper we present our first step in collecting and analyzing memorable spoken quotes. We collect the corpus of memorable quote from TED public speaking and did some preprocessing works, including: 1) match the speech and transcription file, 2) randomly generate the non-memorable corpora, 3) add annotation of popularity factor. The completed corpora consists of memorable and non-memorable quotes in both speech and textual form. The analysis of memorable spoken quotes is done based on speech duration, F0, and popularity. The results reveal that the number of memorable quotes achieve maximum in 15-20 speech duration interval. Analysis on F0 also shows that F0 score of memorable quote corpus apparently is higher than non-memorable quote. This indicates that acoustic may be one of factors which differentiate memorable and non-memorable quotes. As future direction, we will build automatic detection of memorable and popular quotes, which may be learned to enhance spoken dialog system.

[5]Available: http://opensmile.sourceforge.net/

6. ACKNOWLEDGEMENT

Part of this work was supported by JSPS KAKENHI Grant Number 26870371.

7. REFERENCES

[1] R.W. Smith, "Performance measures for the next generation of spoken natural language dialog systems," ISDS, 1997, pp. 37–40.

[2] M. Cavazza, "An empirical study of speech recognition errors in a task-oriented dialogue system," SIGDIAL, 2001, vol. 16, pp. 1–8.

[3] S. Furui, "Recent advances in automatic speech summarization," RIAO, 2007, pp. 90–101.

[4] M. Bulut, S. S. Narayanan, and A. K. Syrdal, "Expressive speech synthesis using a concatenative synthesizer," INTERSPEECH, 2002.

[5] E. Eide, A. Aaron, R. Bakis, W. Hamza, M. Picheny, and J. Pitrelli, "A corpus-based approach to expressive speech synthesis," ISCA Workshop on Speech Synthesis, 2004.

[6] E.T.F. arXiv, "The secret science of memorable quotes," MIT Technol, 2012.

[7] M. Theune, K. Meijs, D. Heylen, and R. Ordelman, "Generating expressive speech for storytelling applications," Audio, Speech, and Language Processing, IEEE Transactions, 2006.

[8] M. Bendersky and D. A. Smith, "A dictionary of wisdom and wit: Learning to extract quotable phrases," NAACL-HLT, 2012, pp. 69–77.

[9] L. Taycher, "Books of the world, stand up and be counted! all 129.864.800 of you," Inside Google blog, 2010.

[10] C. Danescu-Niculescu-Mizil, J. Cheng, J. Kleinberg, and L. Lee, "You had me at hello: How phrasing affects memorability," ACL, 2012, vol. 1, pp. 892–901.

[11] M. Drolet, R. I. Schubotz, and J. Fischer, "Recognizing the authenticity of emotional expressions: F0 contour matters when you need to know," Frontiers, Human neuroscience, 8., 2014.

[12] J. Liscombe, J. Venditti, and J. B. Hirschberg, "Classifying subject ratings of emotional speech using acoustic features," Eurospeech, 2003.

[13] B. Schuller, S. Steidl, and A. Batliner, "The interspeech 2009 emotion challenge," INTERSPEECH, 2009.

[14] F. Eyben, M. Woellmer, and B. Schuller, "opensmile the munich open speech and music interpretation by large space extraction toolkit," Institute for Human-Machine Communication, version 1.0.1, 2010.